FUNDAMENTALS OF ELECTRICITY AND MAGNETISM

International Student Edition

McGraw-Hill Series
in Fundamentals
of Physics

INTRODUCTORY PROGRAM
E. U. Condon, Editor

FUNDAMENTALS OF ELECTRICITY AND MAGNETISM

by Arthur Kip

FUNDAMENTALS OF MODERN PHYSICS

Other titles in preparation

Fundamentals
of Electricity
and Magnetism

ARTHUR F. KIP
Professor of Physics, University of California, Berkeley

Fundamentals of Electricity and Magnetism

INTERNATIONAL STUDENT EDITION

McGRAW-HILL BOOK COMPANY, INC.
New York San Francisco Toronto London
KŌGAKUSHA COMPANY, LTD.
Tokyo

FUNDAMENTALS OF ELECTRICITY
AND MAGNETISM

INTERNATIONAL STUDENT EDITION

II

Library of Congress Catalog Card Number: 61–18259

TOSHO PRINTING CO., LTD. TOKYO, JAPAN

Preface

This is a first course in classical electric and magnetic theory. In such a course there must always be a compromise between the need for presentation of the phenomena of electricity and magnetism and the desire to develop and display the remarkable unity of the theory, at a level consistent with the experience of the beginning student. This compromise has been made on the basis of experience with second-year students in physics, engineering, and chemistry courses at the Massachusetts Institute of Technology and at the University of California.

The basic laws are related to experimental observations, and the theoretical development is connected with experimental phenomena at many points. On the theoretical side, considerable effort has been made to display the remarkable economy of description of the basic phenomena of electromagnetism made possible by the use of Maxwell's equations.

In the development of the subject matter in this course, a knowledge of only elementary calculus and of simple vectors has been assumed. The more complicated concepts required are built

up as the course proceeds. When scalar and vector fields are introduced, care is taken to introduce the use of line and surface integrals and of scalar and cross products. Special sections in the early chapters emphasize the connection between mathematical descriptions and physical problems. Numerous examples are given, both in the body of the text and in selected problems at the ends of chapters, to help the student gain facility in applying mathematical techniques to physical problems.

Since most students come to this course with very little or very cloudy knowledge of all but the simplest phenomena of electricity, practical examples are presented throughout the book. However, since this is a course in physics and not in engineering, practical or useful devices are discussed only to the extent that they help to clarify basic principles. Chapter 13 gives an elementary discussion of the application of electromagnetic theory to the problem of electric discharge in gases and to some simple problems in magnetohydrodynamics.

There are two ways in which this book goes beyond the basic goal of displaying the classical phenomenology and theory of electromagnetism. The first is the inclusion of some of the essential concepts of solid-state physics, where these ideas can aid in the understanding of such fundamental electric phenomena as electric conductivity in metals and semiconductors. Dielectric and magnetic properties of matter are treated in enough detail to prepare the way for later more sophisticated handling in a solid-state course. The second is the introduction, by means of selected topics, of some of the phenomena that illustrate the impact of quantum mechanics on classical electricity and magnetism. In neither case can the treatment of these subjects be comprehensive at this level, but in both cases the student is prepared to appreciate the basic ideas involved. Fuller treatment of quantum-mechanical effects will be given in another volume in this series.

In order to give some of the flavor of the historical development of the subject, a few short excerpts from original papers by Coulomb, Ampère, Faraday, and Maxwell have been included at appropriate places in the text.

Rationalized mks units are used throughout the book, but in Chap. 15, connection is established between these units and the esu and emu systems of units. It seems clear that a single system of units should be used in any introductory course, though it is also

apparent that most students must eventually become familiar with both systems.

The author wishes to express his gratitude to Prof. Alan M. Portis for his many valuable criticisms and suggestions throughout the preparation of this book.

A great debt is also owed to the several classes of students whose reactions to preliminary editions of this book led to many improvements in methods of presentation of the material.

ARTHUR F. KIP

Contents

Introduction

With the great developments in physical science that have occurred in the last thirty years, there has come a need for fundamental revision of the way in which the subject is presented in colleges and universities. In spite of the limited total time available in the curriculum of the four-year course leading to the bachelor's degree in liberal arts and in engineering, there is a growing recognition that fundamental physics requires two years, or at the very least three semesters, for a presentation that is adequate to the needs of those who are going on in any field of science, medicine, or engineering. Soon, it is to be hoped, the one-year without-calculus course, which has been for too long regarded as good enough for liberal arts and premedical students, will be a thing of the past.

Until recently, those who felt a desire to work out a new, modern approach were severely hampered in what they could undertake because of the generally low quality of the instruction in basic science and mathematics that was being offered in most high schools. But this obstacle is rapidly dissolving because of the intensive effort that has gone into the preparation of a better type of

high school physics course through the splendid efforts of the Physical Science Study Committee and others; because of analogous developments in improving the high school course in mathematics, chemistry, and biology; because of the large-scale program of academic-year institutes and summer institutes sponsored by the National Science Foundation for improving the knowledge of high school teachers about the subjects they teach; and because of the development of organized plans for the improvement of science and mathematics teaching at the college level, also sponsored by the National Science Foundation.

Thirty years ago, and even more recently than that, there was a sharp dichotomy among physicists between those primarily interested in research (who were too often deplorably negligent about their teaching) and those primarily interested in teaching (who were too often deplorably negligent about keeping informed on current developments). And it was the teachers who were low on the pecking order of the academic status scale. Fortunately all that is changing now. Outstanding research physicists are devoting energies to improving teaching, and many college teachers are taking a much more active interest in current research progress than formerly.

The new trends, which are still gaining momentum, must be reflected in new approaches in the available textbooks. The present volume, on electricity and magnetism, by Prof Arthur Kip, of the University of California at Berkeley, is the first in a series of four planned as a unit for a modern presentation of physics suitable for men and women who expect to play an effective role, possibly as scientists or engineers, in a world in which genuine understanding of the principles and methods of physics has become so important. It is expected that this book will be used in the second semester of a three- or four-semester sequence, after a course covering mechanics, heat, and the kinetic theory of matter, which will be presented in another volume of this set. Going beyond this volume are two others: one planned for a one-semester course in optics and wave motion and one planned for a one-semester course in atomic and nuclear physics.

Consistently with the improved situation in mathematical instruction that now exists, these books make use of basic concepts of the differential and integral calculus and also of vector methods where needed. Also consistent with the fact that most students

probably will not have delved deeply into these subjects, a minimum use of skill in "wangling" special results by such methods is made. Instead the emphasis is on the development of good physical understanding of physical concepts and the experimental basis underlying their acceptance, rejection, or growth by modification.

It is sincerely hoped that the series will make a genuine contribution toward helping teachers give their students the kind of clear perception of what physics is really about that is so needed for life in the world today.

E. U. CONDON
Chairman, Department of Physics and
Wayman Crow Professor of Physics
Washington University
St. Louis, Missouri

ONE

Electric Charge, Coulomb's Law of Electrostatic Forces

1.1 Introduction

A thorough investigation of the behavior of *electric charges* (or *electrostatic charges*) will lead us to the complete theory of electromagnetism, and although present theory, as we shall see, tends to place most emphasis on rather abstract quantities such as electric fields, potential, and lines of force, it is the demonstrable reality of electric charges that forms the basis of all our ideas concerning electromagnetism.

In our treatment of electric charges we start with the simplest experimental facts and indicate how the present framework of ideas and methods of theoretical treatment has evolved. For simplicity we consider at first the idealized situation as it would occur if the charges were in a vacuum. Later we shall study the usually small perturbing effects of air and other matter on our simple results. The treatment of charges at rest will first concern us, and only much later shall we consider the effects of moving charges, i.e., currents. The study of forces on currents will be related to the

phenomena of magnetism. Finally, we shall show that the propagation of energy by electromagnetic waves, as in, for example, radio waves and light, is to be understood on the basis of the phenomena we have already studied.

The knowledge of the existence of electrostatic charge goes back at least as far as the time of the ancient Greeks, around 600 B.C. We can repeat the observations of the Greeks by rubbing a rod of amber or hard rubber with a piece of fur. After this it will be found that small bits of paper or other light materials are attracted to the rod. No particular advance was made in the understanding of this phenomenon until about 1600, when William Gilbert, court physician to Queen Elizabeth, began a detailed study of the kinds of materials that would behave as amber. These he described as *electric* (from the Greek word for amber, *elektron*). Materials that Gilbert found unable to show this attractive force he called *nonelectrics*. We now call these two kinds of materials *insulators* and *conductors*.

The next important step in the development of ideas about charges came about 100 years later. Du Fay showed that there are two kinds of electrification. By rubbing various kinds of insulators together, he was able to show that under some conditions they repel each other. His results could be explained by postulating two kinds of charge. Forces between bodies having like charge are found to be repulsive, while forces between unlike charges are attractive. The quantitative theory assigns a plus sign to one type of charge and a minus sign to the other, as was first suggested by Benjamin Franklin. Which sign is given to which kind of charge is arbitrary (and unimportant), but as we shall see, a sign convention allows us to make a very concise mathematical formulation of the experimental facts.

We now leave the qualitative discussion of electric charge and begin the study in quantitative form.

1.2 Electric Charge, Coulomb's Law of Force

It is remarkable that all the simple phenomena involving charges at rest can be described very well by the equation

$$\mathbf{F} \propto \sum_i \frac{q_i q'}{r_i^2} \qquad \text{(vector sum)} \tag{1.1}$$

This is one form of the law of electrostatic forces. We owe its formulation to experiments by Priestley in 1767 that were repeated by Coulomb in 1785. Usually called Coulomb's law of force, it expresses the force **F** between point charges q' and q_1, separated by a distance r_1.

Although this is not the most general description possible, it is a reasonably complete statement of the ideas about charges. Since we are at the moment most interested in the physical ideas regarding charges, we shall not discuss Eq. (1.1) now but, instead, shall write down explicitly the facts of interest about charges and then show how the remarkable shorthand description above does indeed include all the ideas of interest about stationary charge. Our discussion at present is limited to situations that involve *point* charges; that is, we shall consider only charges that are concentrated in a region whose dimensions are small compared with the distance to other charges.

The following statements are derived directly or indirectly from experiment, though we shall not go into details in all cases. They not only give the known facts but also define the point of view from which we explain the experimental facts today.

1. The presence of electric charges is made known by the existence of attractive or repulsive forces between them. These forces can be large enough in the laboratory to allow quantitative measurements to be made.

2. There are two kinds of electric charge. The force between *like* charges is *repulsive* and acts along the line joining them. Between *unlike* charges the force is *attractive* and also acts along the line joining the charges, which are named *positive* and *negative* (quite arbitrarily).

3. The force between a given pair of charges is inversely proportional to the square of the distance between them.

4. The force between two charges is proportional to the *quantity* of one charge multiplied by the *quantity* of the other.

5. The force between any two charges is independent of the presence of other charges.

We shall now examine these statements in some detail. Statement 1 is well illustrated by the early experiments with amber and bits of paper or other material as performed by the Greeks. Statement 2 is a little more sophisticated. The experimental proof requires first that we have a method of producing both kinds of

charge at will and then that we be able to check the direction of the force in the several cases. We shall first describe a way of showing that like charges repel each other. Suppose we rub a rod of amber or hard rubber with wool or cat's fur. The effect of this, it turns out, is to leave an excess of negative charges on the rod and an equivalent positive charge on the wool. Because like charges repel, as we shall soon show, the excess negative charge will tend to leak off onto any object to which the rod is touched. If we then touch the rod to a small ball of paper or other material, suspended on an insulating thread (to prevent charge from leaking off), some of the excess charge will be transferred to the ball. We may then repeat the experiment using another ball of paper, so we now have two balls with the same kind of charge, and the first crucial experiment can be performed. The question is whether the two balls attract or repel each other, and we find that they indeed repel.

If an exactly similar experiment is now performed, this time using, for example, a glass rod rubbed with silk to produce positively charged balls, we shall again find repulsive forces between the charges. The final experiment will involve the use of one ball charged with the amber rod and one charged with the glass rod. In this case, we find an attractive force, in contrast to the two earlier experiments. Further experiments are necessary to demonstrate the reproducibility of the effects and to show that all charges can be placed in either one or the other category, but the above experiments are crucial ones for determining the sign of forces, as well as for determining that they act along the line joining the charges.

Statements 3 and 4 are the first ones requiring quantitative measurements for their verification. A possible, though perhaps

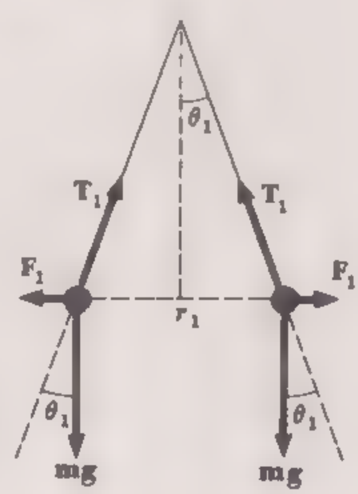


***Fig. 1.1** Arrangement for measuring force of repulsion between two equally charged pith balls.*

not very practical, way to verify statement 3 is given next. Later a much more satisfying argument will be given.

A demonstration of the validity of the inverse-square law of repulsion between like charges. Suppose we have like charges of equal magnitude on two pith balls of equal mass, suspended in equilibrium on insulating threads of negligible mass as shown in Fig. 1.1. The equal charge and mass requirement is not necessary, but is made to simplify the calculation. Since the resultant of the electrostatic force **F** and the gravitational force **mg** must be equal and opposite to the tension force **T**,

$$\frac{F_1}{mg} = \tan \theta_1$$

When we shorten the suspending threads, the system takes up a new equilibrium position, with a new angle θ_2. The new repulsive electrostatic force will be given by $F_2/mg = \tan \theta_2$. Thus by measuring θ_1 and θ_2, we can get the ratio F_1/F_2 since **mg** does not change. If we also measure the separations between the charges, r_1 and r_2, we can relate the ratio of forces to the ratio of separations. We shall find experimentally that $\frac{F_1}{F_2} = \frac{r_2{}^2}{r_1{}^2}$. This result is of course consistent with statement 3, that

$$F \propto \frac{1}{r^2} \tag{1.2}$$

It is of interest to quote briefly from the original papers of Coulomb[1] part of his description of his quantitative measurements of the law of forces between charged particles. Coulomb had determined the law of torsion of wires, after which he applied the torsion method to the investigation of electrostatic forces. His torsion balance was arranged so that the force between two charged bodies resulted in a twisting of a fine suspension wire. He compared the amount of twisting for various separations between the charged bodies and from this was able to induce the inverse-square law:

> In a memoir presented to the Academy in 1784, I determined by experiment the laws of force of torsion of a metallic wire. . . .

[1] Charles Augustin de Coulomb, Memoires sur l'électricité et le magnétisme, *Mem. Acad. Roy. Sci.*, pp. 569*ff*. (1788). The quotation is from the 1785 volume, published in 1788—there were delays in publishing in those days too!

I showed in the same memoir that by using this force of torsion it was possible to measure with precision very small forces, as for example, a ten thousandth of a grain. . . .

I submit today to the Academy an electric balance constructed on the same principle; it measures very exactly the state and the electric force of a body however slightly it is charged. . . .

. . . we go on to give the method which we have used to determine the fundamental law according to which electrified bodies repel each other.

In the third trial the suspension wire was twisted through 567 degrees and the two balls are separated by only 8 degrees and a half. The total torsion was consequently 576 degrees, four times that of the second trial, and the distance of the two balls in this third trial lacked only one-half degree of being reduced to half of that at which it stood in the second trial. It results then from these three trials that the repulsive action which the two balls exert on each other when they are electrified similarly is in the inverse ratio of the square of the distances.

Statement 4 combines the law of interaction between charges with the assumption of a kind of charge conservation. That is, it implies that we can keep track of the amount of charge we are dealing with and that we can then relate the magnitude of forces between charges to the *quantities* of charges involved. The idea is stated much more easily in mathematical terms as follows: Two charges at a fixed distance apart exert forces on each other such that the force

$$F \propto q_1 q_2 \tag{1.3}$$

where q_1 and q_2 measure the quantity of charge at each position. We can think of an experiment involving the measurement of pairs of charges by the amount of force exerted between them at given distances and the subsequent combining of some of the charges in order to test both the additivity (or conservation) of charges and the multiplicative law of forces as given above. The experimental results are indeed in agreement with these two ideas. Charge conservation here includes the effects of the existence of two kinds of charge, and we can now see the beautiful simplicity resulting from the use of + and − terminology for them. The terminology exactly fits the fact that equal amounts of + and − charges cancel, whereas similar charges add. As regards the law of force, we need make only the one convention in the equation

$F \propto q_1q_2$ that a positive force shall mean a force of repulsion, whereupon we find that the equation gives the right kind of force for any combination of pairs of forces. There would be no change in the predicted direction of forces if we interchanged the names of the two kinds of charge.

It is now possible to combine the last two equations [(1.2) and (1.3)] into a single equation:

$$F \propto \frac{q_1q_2}{r^2} \tag{1.4}$$

It is certainly very satisfying that so many ideas can be so simply contained in one equation. Actually there is still another idea, which we have not discussed, which is already implied by the mathematical terminology. Since a force is a vector quantity, Eq. (1.4) is a vector equation. Mathematically this equation says that if we choose to look at the force on one charge, say q_1, it will be the vector sum of the forces due to any number of other charges q_2, q_3, etc., at the distances r_{12}, r_{13}, etc. Thus we would write

$$\mathbf{F} \propto \left(\frac{q_1q_2}{{r_{12}}^2} + \frac{q_1q_3}{{r_{13}}^2} + \cdots\right) \text{(vector sum)}$$

Physically this idea will be correct only if the force between two charges is independent of the presence of any other charges. But this is exactly the experimental fact given in statement 5, so we can accept the mathematical formulation as a complete description of the experimental facts.

Vector notation It is useful to introduce one more kind of mathematical terminology that will aid in the description of the physical situation. We have noted that (1.4) is a vector equation, since it involves force, which is a vector quantity. In order to indicate that the right-hand side of the equation is also a vector quantity, we shall introduce the *unit vector* $\hat{\mathbf{r}}$, which has the direction of the vector (in this case, the direction of the radius vector $\mathbf{r}$ from q_2 to q_1) and a magnitude of unity. Thus we rewrite Eq. (1.4):

$$\mathbf{F}_1 \propto \frac{q_1q_2}{r^2}\,\hat{\mathbf{r}} \tag{1.4a}$$

There is no change in the meaning of the equation, but we have made more explicit the vector nature of both sides. Similarly, the equation that follows (1.4) becomes

$$\mathbf{F}_1 \propto \left(\frac{q_1q_2}{{r_{12}}^2}\,\hat{\mathbf{r}}_{12} + \frac{q_1q_3}{{r_{13}}^2}\,\hat{\mathbf{r}}_{13} + \cdots\right)$$

Here the unit vectors $\hat{\mathbf{r}}_{12}$ and $\hat{\mathbf{r}}_{13}$ have the directions of the lines from q_2 to q_1 and q_3 to q_1, respectively. This relieves us of the necessity of writing *vector sums* after the equation, since the vector nature of the sum is contained in the mathematical description.[1]

We have now only one more task to replace the proportionality sign in the equation with the more explicit equals sign. The usual method is to write the equation with a proportionality factor, which must then be evaluated. Thus our equation

$$F \propto \frac{q_1 q_2}{r^2}\,\hat{\mathbf{r}} \qquad \text{becomes} \qquad F = K\frac{q_1 q_2}{r^2}\,\hat{\mathbf{r}}$$

where K is the proportionality factor. The value of K depends on the choice of units for F, r, and q. If the units of F and r have already been adopted from a study of mechanics (as F in newtons and r in meters in the mks system), the value of K then depends solely on the choice of the unit of q. In the mks system we define the unit of q from totally different considerations based on magnetic forces between electric currents; the unit is called the *coulomb*. In that case K becomes a quantity that has to be measured experimentally. For reasons that become clear later, it is customary to write K as $(4\pi\epsilon_0)^{-1}$, so the equation becomes

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2}\,\hat{\mathbf{r}} \tag{1.5}$$

Thus defined, the q's are measured in coulombs, and ϵ_0, called the *permittivity* of free space, is found by experiment to have the value 8.85×10^{-12} coulomb^2/newton-m^2. It is convenient to remember the equivalent relationship $1/4\pi\epsilon_0 = 9 \times 10^9$ newton-m^2/coulomb^2. The fundamental definition of the coulomb comes from the magnetic effect of moving charges, discussed in Sec. 6.4. The coulomb is the unit of charge in both the mks system of units and the *practical* system (see Chap. 15).

[1] Another common way of handling the vector nature of equations like (1.4*a*) is to write it as

$$\mathbf{F}_1 \propto \frac{q_1 q_2 \mathbf{r}}{r^3} \tag{1.4b}$$

where $\mathbf{r}$ is now a vector with the magnitude of the distance r. The r^2 in the denominator has been changed to r^3 to compensate for this, so the meaning is identical with that of Eq. (1.4*a*). We shall use the unit-vector notation.

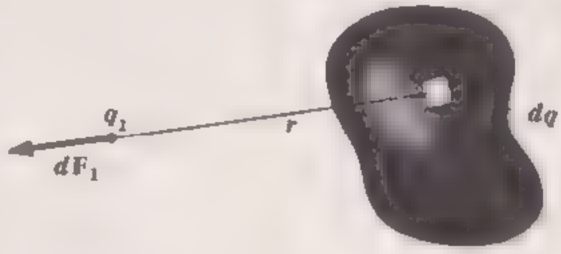
For the force on q_1 due to a number of charges, we may write

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_1 q_i}{r_i^2} \hat{\mathbf{r}}_i \tag{1.6}$$

which amounts to the same thing as Eq. (1.1). A true understanding of this equation gives insight into a great deal of the physics involved in the concept of charges and of the forces between them.

One slight modification allows us to handle situations where the charge is spread over a region instead of being concentrated at particular points. Suppose we have a point charge q_1, which is near a region of continuous charge distribution, as indicated in Fig. 1.2,

Fig. 1.2 Calculation of the force on a charge q_1 due to a continuous distribution of charge. The fields due to each element dq at its distance r must be added vectorially. $\hat{\mathbf{r}}$ points from dq to q_1, and $d\mathbf{F}_1$ is the contribution of the interaction between q and dq to $\mathbf{F}_1$.



where we wish to calculate the total force on q_1 due to all the other charge. As suggested by our discussion so far, we shall need to make a vector sum of all the forces due to all the small charges dq distributed over the region. Thus we would write

$$\mathbf{F}_1 = \frac{1}{4\pi\epsilon_0} \int \frac{q_1\,dq}{r^2} \hat{\mathbf{r}} = \frac{q_1}{4\pi\epsilon_0} \int \frac{dq}{r^2} \hat{\mathbf{r}} \tag{1.7}$$

where $\hat{\mathbf{r}}$ is a variable unit vector that points from each dq toward the location of the charge q_1.

1.3 Examples

We give below a few examples of particular situations involving the ideas discussed so far. Certain parts of the work are dealt with only briefly, since more general methods will be applied to them in following chapters.

a Force between two point charges Two point charges q_1 and q_2 are separated by a distance r, as shown in Fig. 1.3. Find the force acting between them. This is clearly the most elementary case

***Fig. 1.3** Two point charges separated by a distance r.*

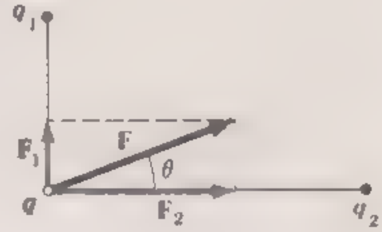
possible. In the mks system, we must have q_1 and q_2 in coulombs, r in meters, and F in newtons if we use

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{q_1q_2}{r^2}\hat{\mathbf{r}} = 9 \times 10^9 \frac{q_1q_2}{r^2} \qquad \text{newtons}$$

To define the result completely, we must state that the forces act along the line between the charges and are either attractive (negative) or repulsive (positive).

b Force on one charge due to two others For convenience we choose a simple right-angle geometry as shown in Fig. 1.4. This

***Fig. 1.4** Resultant force on q due to q_1 and q_2 obtained by vector addition of individual forces.*



solution follows immediately if we remember the vector nature of force (and the implied independence of the force between two charges of the presence of other charges). We calculate $\mathbf{F}_1$ and $\mathbf{F}_2$, the forces on q, independently according to Coulomb's law, as in Example 1.3*a*, and then perform the *vector* addition. In this simple case, $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$ becomes

$$\mathbf{F} = \sqrt{|\mathbf{F}_1|^2 + |\mathbf{F}_2|^2}$$

The direction of $\mathbf{F}$ is given of course by $F_1/F_2 = \tan\theta$ or $\theta = \tan^{-1} F_1/F_2$.

c Force due to linear charge distribution Imagine a long, thin stick (see Fig. 1.5) with a uniform distribution of excess charge on it. Suppose the total excess charge on the stick is Q. What will be the force of these charges on a charge q at a distance a from the

stick along a line through the stick? This is a case requiring the use of an equation like (1.7), so that we can integrate over the entire charge distribution. Thus we must find an expression that allows us to sum up each differential piece of Q, keeping track of its distance

Fig. 1.5 Calculation of force on a charge q due to a continuous linear distribution of charge.

from q. A convenient way is to establish a representative element of charge dQ at a distance x from q. The force on q due to this element will be

$$dF = \frac{q}{4\pi\epsilon_0}\frac{dQ}{x^2}$$

In order to integrate this, we must somehow relate each dQ to its appropriate x. We see at once that the *linear density* of charge along the stick is Q/L, which we might call μ. Then the amount of charge in our element is $\mu\, dx$ if its length is dx. Our equation is now

$$dF = \frac{q\mu}{4\pi\epsilon_0}\frac{dx}{x^2}$$

which can be integrated. Thus,

$$F = \int_a^{L+a} \frac{qQ}{4\pi\epsilon_0 L}\frac{dx}{x^2} = \frac{qQ}{4\pi\epsilon_0 L}\int_a^{L+a}\frac{dx}{x^2} = \frac{qQ}{4\pi\epsilon_0 L}\left(\frac{1}{a} - \frac{1}{L+a}\right)$$

This simplifies to

$$F = +\frac{qQ}{4\pi\epsilon_0}\frac{1}{a(L+a)} \qquad \text{newtons}$$

The positive sign indicates that the force is repulsive when q and Q have the same sign. It is easy to show that this answer is reasonable. Suppose we let $a \gg L$; then the forces should approximate those between two point charges q and Q. But this is just our result if we neglect L, as would be justified if $a \gg L$. That is, the further away we get from the stick the more nearly it acts like a point charge.

Another comment should be made about this problem: We have chosen a particularly simple example in which the direction

of the force from each element of charge is the same. This allows us to neglect the vector nature of the integration since the sum of a number of vectors all pointing in the same direction is just the arithmetic sum of their magnitudes. In some later problems we shall study the more general case.

1.4 Electric Charge and Matter

Electric charges and matter are intimately connected. A short discussion of a few of the basic ideas will be helpful later when we describe some of the experiments on which our understanding of electricity is based.

Electric charge is basic in the structure of atoms. Around a positively charged *nucleus* are situated the negatively charged *electrons*. One of the important forces in atoms is the attraction between the oppositely charged electrons and nuclei. If large enough forces are applied to electrons associated with atoms, some of them may be removed despite the attractive force of the nucleus. This is the source of electrons which make up the electric current which plays such a dominant role in electricity and magnetism.

When atoms combine to form solids, it often happens that one or more of the electrons normally bound to each atom are liberated and can wander around more or less freely in the material. These are the *conduction electrons* in metals. When such freeing of the electrons does not occur, we speak of the material as an insulator or dielectric. In such materials the electrons are not free to move around, and thus external forces cannot produce currents within the material. We shall, however, discuss electrical effects in these materials later, when we shall see that since the electric-charge centers are not tied down rigidly, their slight motion as they stretch their bonds under external forces causes a host of interesting and revealing dielectric effects.

We could establish a third category of solid in which it is possible to free some of the bound charges and thus allow current to flow within the matter, at least as long as the charge carriers remain free. Such a material is called a *semiconductor*. Bound charges can be freed by thermal vibrations within a semiconductor, by incident light, or by the application of externally produced electric fields.

The freeing of electrons from atoms at the surface of solids accounts for the charging of bodies by rubbing materials like amber and fur together. Of two different materials rubbed together, the one from which electrons are most easily separated tends to lose electrons and get a positive charge, and the other one becomes negatively charged by collecting excess electrons.

1.5 Some Relevant Apparatus

The electroscope One of the early instruments for measuring the presence of charge is the gold-leaf electroscope. Two strips of gold foil are connected to an insulated metal rod as shown in Fig. 1.6.

Fig. 1.6 Sketch of a gold-leaf electroscope.

If an excess charge is placed on the rod, part of the charge flows to the foils, which are then repelled from each other. The extent of the separation of the foils will depend quantitatively on the amount of charge on the electroscope. This rather primitive indicator of charge is still used in demonstration experiments.

The electrophorus This early device is one of the simplest for obtaining charge separation. A modern form of the apparatus,

Fig. 1.7 Electrophorus apparatus. Induced charge produced by charge held on insulating plate A can be collected on insulated metal plate B.

often used for demonstration purposes, is shown in Fig. 1.7. Plate A is an appropriate insulator such as sealing wax, which can be charged on its surface by rubbing with cat's fur. B is a metal plate,

held by means of an insulating handle. The object of the device is to charge the metal plate repeatedly by a suitable cycle of operations. After each charging of the plate, all or most of its excess charge can be placed on some insulated body, such as an electroscope. The method of operation is as follows: After the charge is placed on the surface of plate A, plate B is brought up fairly close. Under the influence of the negative charges on A, some of the mobile negative charges in B are repelled away from its bottom surface, leaving it with a net positive charge. An equal excess negative charge collects on the top surface, since the metal plate is originally neutral. At this time, the top of the metal plate is touched with a finger. Since the human body is a fairly good conductor, this action allows some of the excess electrons on top of the plate to flow off through the finger under the forces of mutual repulsion between electrons. Plate B now has an excess positive charge that can be transferred to any apparatus as desired. (Actually, electrons will flow to the metal plate, neutralizing it and leaving the apparatus with a net positive charge, but we may perfectly well think of this as a flow of positive charge from the plate to the apparatus.) The process does not modify the excess charge originally placed on the insulating plate A. Therefore the foregoing cycle may be repeated many times with a resulting large collection of charge on the apparatus. This method of charging the metal plate is known as charging by induction.

Attraction of neutral bodies by charged bodies As mentioned earlier, small bits of paper are found to be attracted to a charged rod. Since the paper is normally uncharged, this is at first surprising. It is easily explained when we realize that although the total charge on the paper is zero, under the influence of the charged rod the equal positive and negative charges on the paper will redistrib-

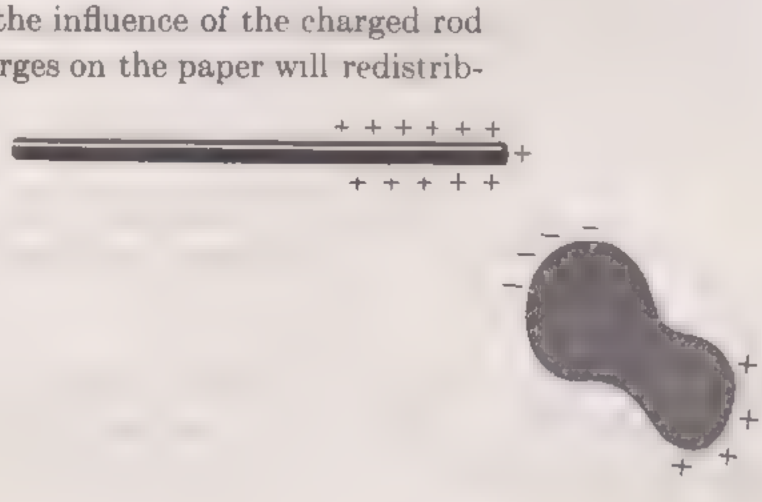


Fig. 1.8 Attraction of a neutral body by a charged rod.

ute themselves much as they do on the metal plate of the electrophorus apparatus; thus a net attractive force is produced, as shown in Fig. 1.8. Such redistribution of charge, we shall see later, can take place even in an insulator.

If the bits of paper touch the charged rod, they tend to fly off at once. This results from the leaking of some of the charge on the rod onto the paper, thus bringing into play the mutual repulsion between the like charges on the rod and paper.

Electrostatic generators Two types of electrostatic generators are often used for obtaining large concentrations of charge, at resultant high *voltage* or high *potential* (these terms are discussed later on). One, the Wimshurst machine, is of early origin and is a complicated machine involving rotating plates which are charged by induction and from which the collected charges are removed by metal conductors. A more effective device is the Van de Graaff generator. This is a modern machine, large models of which can produce potentials of millions of volts for use in accelerating charged particles for nuclear experiments. It consists of an insulating belt onto one end of which charges are sprayed; these are carried to the other end and collected. Small demonstration models of the Van de Graaff generator are often seen in physics laboratories.

PROBLEMS

1.1 Three point charges $+Q_1$, $-Q_2$, and $+Q_3$ are equally spaced along a line as shown in Fig. P1.1. If the magnitudes of Q_1 and Q_2 are equal, what must be the magnitude of Q_3 in order that the net force on Q_1 be zero?

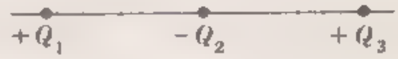


Fig. P 1.1

1.2 Three identical point charges of Q coulombs are placed at the vertices of an equilateral triangle, 10 cm apart. Calculate the force on each charge.

1.3 *a* Find the force on a point charge of $2Q$ coulombs at the center of a square 20 cm on a side if four identical point charges of Q coulombs are located at the corners of the square.

b Find the force on the charge at the center of the square when one of the corner charges is removed.

1.4 A point charge of Q_2 coulombs is located on the x axis a distance a meters from another point charge of Q_1 coulombs, as shown in Fig. P1.4. Calculate the force on Q_2 and then calculate the work to move Q_2 from a to a distance b from Q_1.

Fig. P 1.4

1.5 A thin circular ring of 3 cm radius has a total charge of 10^{-3} coulomb uniformly distributed on it. What is the force on a charge of 10^{-2} coulomb at its center? What would be the force on this charge if it were placed at a distance of 4 cm from the ring, along its axis?

1.6 Two charges of Q coulombs each are placed at two opposite corners of a square. What additional charges q placed at each of the other two corners will reduce the resultant electric force on each of the charges Q to zero? Is it possible to choose these charges so that the resultant force on *all* the charges is zero?

1.7 Compute the ratio between the electrostatic repulsion and the gravitational attraction between two electrons. The charge on an electron is -1.6×10^{-19} coulomb and its mass is 9.0×10^{-31} kg. The gravitational constant is 6.670×10^{-11} newton-m²/kg-m².

1.8 In a hydrogen atom the negative electron moves in an orbit around the (much heavier) positive proton, bound by the attractive coulomb force. Assuming that the orbit is circular and has a radius of 0.528×10^{-8} cm, calculate the number of revolutions per second made by the electron; calculate the angular momentum of the system. How large would the hydrogen atom be if it moved with the same angular momentum but was bound by gravitational attraction?

TWO

The Electric Field

2.1 *Introduction*

In this chapter we develop a somewhat more general way of handling the kinds of force problems discussed in Chap. 1. This involves use of the *electric field*. This new point of view allows us to extend our ideas greatly and to simplify our understanding of charges and the forces between them. Following the introduction of the electric field, we discuss the use of lines of force; finally, we show how *Gauss' law* can be developed and used to great advantage in certain situations.

2.2 *The Electric Field*

The electric field at a given point in space can be defined as the force per unit positive charge that would act on such a charge were it placed at that point. Since it measures a force, it must be a vector quantity. We shall call it **E**. Thus when we place a test charge q at

a given point in space, the force on it due to the other charges in the region is given by

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0} \sum_i \frac{qq_i}{r_i^2} \hat{\mathbf{r}}_i \tag{1.6}$$

If the point of interest is at a in Fig. 2.1 and if the fixed charges

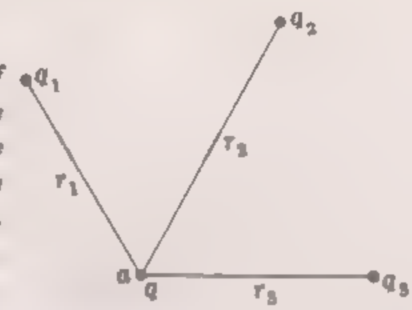


Fig. 2.1 *Calculation of force on charge q due to charges q_1, q_2, and q_3. The electric field at a is the force per unit charge at a due to the charges q_1, q_2, and q_3.*

q_1, q_2, and q_3 are the only ones causing forces on the test charge q, the force on it is calculated from Eq. (1.6) as

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\left(\frac{qq_1}{r_1^2}\hat{\mathbf{r}}_1 + \frac{qq_2}{r_2^2}\hat{\mathbf{r}}_2 + \frac{qq_3}{r_3^2}\hat{\mathbf{r}}_3\right)$$

Now, to assume the new viewpoint, the electric field $\mathbf{E}$ at position a will be given by

$$\mathbf{E} = \frac{\mathbf{F}}{q} = \frac{1}{4\pi\epsilon_0}\left(\frac{q_1}{r_1^2}\hat{\mathbf{r}}_1 + \frac{q_2}{r_2^2}\hat{\mathbf{r}}_2 + \frac{q_3}{r_3^2}\hat{\mathbf{r}}_3\right)$$

That is, we have merely divided through by q. Our calculation now gives the force not on q but on a *unit* charge at the point a, which is just as we defined $\mathbf{E}$, the electric field at a due to the charges q_1, q_2, and q_3 (but not q).

In general, the field at any given point is given by

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{r_i^2} \hat{\mathbf{r}} \tag{2.1}$$

or, for a continuous charge distribution, by

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \int \frac{dq}{r^2} \hat{\mathbf{r}} \tag{2.2}$$

where $\hat{\mathbf{r}}$ is the unit vector of variable direction pointing from q_i or dq toward the point in space for which $\mathbf{E}$ is being calculated.

This new quantity may now be used to calculate the force on a certain charge q_0 at any particular place. The procedure is simply to calculate the field **E** at the point in question, and then

$$\mathbf{F} = q_0\mathbf{E} \tag{2.3}$$

In general, the force on any charge is obtainable by calculating the electric field and then multiplying this by the magnitude of the charge. Since Eq. (2.3) is a vector equation, **E** and **F** are along the same line and are in the same sense if q_0 is positive, but oppositely directed if q_0 is negative. For this simple example we could as well have calculated the force on our charge q_0 directly from Eq. (1.6) without reference to the electric field.

Introduction of the field **E** allows a more general aproach, which is important in understanding much more complicated situations. Thus we have now introduced a *vector field*. We can imagine a plot of the electric field in a region as a diagram like that in Fig. 2.2 which would tell us the magnitude and direction

***Fig. 2.2** The electric field is a vector field. To each point in space may be assigned a magnitude and direction.*

of the force per unit charge at every position in the region. We might think of the field as describing the condition of space in a given region. We are thus deemphasizing the individual charges that cause the field and are instead thinking more about the effect of their presence on the space around them. We shall have much more to do with this vector field further on. Here we examine only a few simple problems. It is interesting to note that the value of the electric field at particular points may well be zero. Also, the value of this new vector quantity is always finite (or zero) as long as we stay away from the immediate vicinity of any of the charges causing the field. If we get too close to any of these, **E** tends to infinity, since the distance r in q/r^2 tends to zero. In a major part of our study of electricity we avoid this difficulty by staying suffi-

ciently far from electric charges. Since they are actually very highly localized—on electrons and protons, for example—this is easy to do without interfering with the usefulness of our study.

We now point out a necessary elaboration on our earlier definition of electric field. More accurately, we define the field thus:

$$\mathbf{E} = \lim_{\Delta q \to 0} \frac{\Delta \mathbf{F}}{\Delta q} \qquad \text{or} \qquad \mathbf{E} = \frac{d\mathbf{F}}{dq} \tag{2.4}$$

That is, the field is the limiting value of the force per unit charge as the test charge is made vanishingly small. This elaboration is needed because there are situations in which a finite test charge, by its presence, alters the distribution of charges causing the field. Suppose, for example, that the electric field is due to charges placed on conductors. When we bring up a finite charge, the other charges producing the field redistribute themselves as a result of the forces between them and the test charge. The field is thus modified by the presence of the test charge according to its size. The definition of Eq. (2.4) avoids this pitfall.

Equations (2.1) and (2.2) are extremely important. They provide the method by which it is always possible to calculate the electric field at a given point due to a distribution of charges. We next show some examples of field calculation according to this scheme.

2.3 Examples, Calculation of Electric Fields

a Field of an array of point charges Given point charges q_1 and q_2 as shown in Fig. 2.3 at distances a and b from the origin on the x and

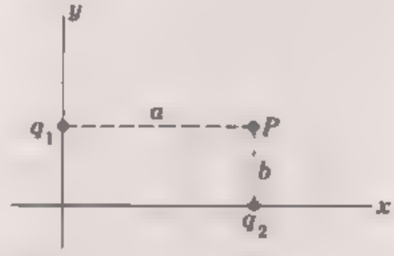


Fig. 2.3 *Calculation of the field at P due to charges q_1 and q_2.*

y axes, find the value of $\mathbf{E}$ at the position P (a,b). Using Eq. (2.1), we write

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\left(\frac{q_1}{a^2}\hat{\mathbf{r}}_1 + \frac{q_2}{b^2}\hat{\mathbf{r}}_2\right)$$

Since for this case the two vectors are at right angles, the calculation of the field at P becomes

$$E = \frac{1}{4\pi\epsilon_0}\sqrt{\left(\frac{q_1}{a^2}\right)^2 + \left(\frac{q_2}{b^2}\right)^2}$$

and the angle θ, say, between the resultant field and the x axis is found by using

$$\frac{q_2\ b^2}{q_1/a^2} = \tan\theta$$

b Field out from a long uniformly charged rod Contrast this with Example 1.3c, from which it differs by asking for the electric field rather than for the force on a certain charge. Also, in the present example the point of interest is to the side of the line containing the charge distribution. This brings out the full complexity of the vector summation in the integral. As shown in Fig. 2.4, we shall

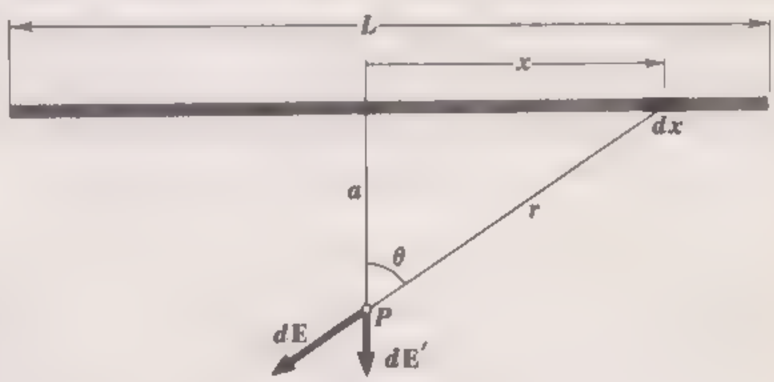


***Fig. 2.4** Calculation of the field due to a uniform linear array of charges.*

calculate the electric field **E** at the point P, a distance a along the perpendicular bisector of the rod. We let the linear charge density be μ as before. It is Q/L, where Q is the total charge on the rod.

The problem is simplified if we replace the variable x by the variable angle θ. We may then express the variable distances x and r in terms of θ and the fixed distance a.

The component of the field at P due to the element of charge $\mu\,dx$ is $dE = \dfrac{1}{4\pi\epsilon_0}\dfrac{\mu\,dx}{r^2}$ by Coulomb's law. Since $x/a = \tan\theta$ and $a/r = \cos\theta$, we have $dx = a\sec^2\theta\,d\theta$ and $r = a/\cos\theta$. Substitution then gives $dE = (1/4\pi\epsilon_0)\ d\theta$. Because the various components $d\mathbf{E}$ from the various charge elements are not in the same direction, they must be added vectorially. The problem of taking the vector

sum of these individual components turns out to be trivial when we use the obvious symmetry of the problem. That is, if the point P is opposite the center of the uniformly charged rod, the x components at P of each $d\mathbf{E}$ will cancel (since for each element at x there will be an equivalent at $-x$), whereas the y components along a will add. We therefore take the sum of the y components of each $d\mathbf{E}$ to obtain the required vector sum. We call such a component $d\mathbf{E}'$, given by

$$dE' = \frac{1}{4\pi\epsilon_0}\frac{\mu}{a}\cos\theta\,d\theta$$

The total field $\mathbf{E}'$ at P for a very long rod is then obtained from

$$E' = \frac{2}{4\pi\epsilon_0}\frac{\mu}{a}\int_0^{\pi/2}\cos\theta\,d\theta = \frac{2\mu}{4\pi\epsilon_0 a}$$

If the rod is not very long, appropriate changes are necessary in the upper limit of integration. For a finite rod and for a point P not opposite the middle of the rod, we cannot use the simple symmetry property, and the solution is more difficult. Coulomb's $1/r^2$ law refers only to point charges. In this problem, for example, the field falls off as we move away from a long rod as $1/a$.

c Field due to an electric dipole A pair of equal and opposite point charges separated by a vector distance $\mathbf{a}$ is called a *dipole.* The vector $\mathbf{a}$ is drawn from the negative to the positive charge and is along the *axis* of the dipole. Calculation of the field of a dipole is

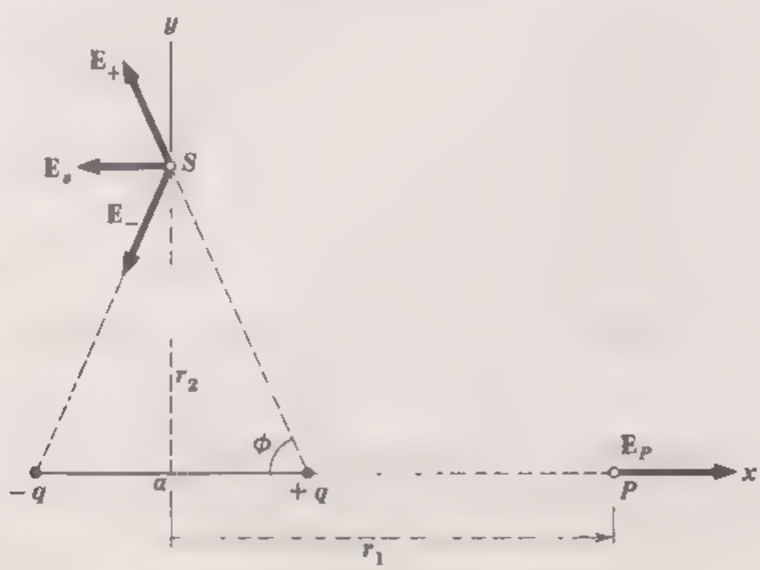


***Fig. 2.5** Calculation of field of a dipole of moment* $p = q\mathbf{a}$.

an easy problem that does not involve integration because only point charges are concerned. It is an important problem because of the common occurrence of dipoles. A molecule made up of a positive and a negative ion is one example of an electric dipole in nature. Also, the dipole is often the most convenient first step in the description of more complicated arrays of charge.

We calculate the field of the dipole along its axis and perpendicular to the axis from the center of the dipole. Later we make the calculation for a general position. For convenience, we place the dipole along the x axis at the origin of the coordinate system as shown in Fig. 2.5. By symmetry, results along the y and z axes will be identical, so we limit our discussion to the x and y axes.

We first discuss the field at P at a distance r_1 from the center of the dipole. This is given by

$$E_P = \frac{1}{4\pi\epsilon_0}\left[\frac{q}{(r_1 - a/2)^2} - \frac{q}{(r_1 + a/2)^2}\right]$$

$$= \frac{q}{4\pi\epsilon_0}\left[\frac{2r_1 a}{(r_1^2 - a^2/4)^2}\right]$$

At S the fields due to $+q$ and $-q$ are designated as $\mathbf{E}_+$ and $\mathbf{E}_-$, respectively. Their y components cancel, while their x components add to yield the resulting field $\mathbf{E}_s$. Thus,

$$E_S = |\mathbf{E}_+| \cos\theta + |\mathbf{E}_-| \cos\theta$$

or

$$E_S = \frac{1}{4\pi\epsilon_0}\left[\frac{q}{r_2^2 + (a/2)^2} + \frac{q}{r_2^2 + (a/2)^2}\right] \frac{a/2}{[r_2^2 + (a/2)^2]^{1/2}}$$

$$= \frac{q}{4\pi\epsilon_0} \frac{a}{[r_2^2 + (a/2)^2]^{3/2}}$$

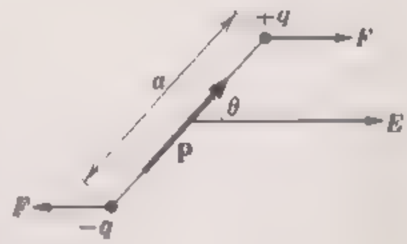
A very useful approximation valid at distances much greater than the separation a results from neglecting $a^2/4$ compared with r^2. Then we find

$$E_P = \frac{1}{4\pi\epsilon_0} \frac{2p}{r_1^3} \tag{2.5}$$

$$E_S = \frac{1}{4\pi\epsilon_0} \frac{p}{r_2^3} \tag{2.6}$$

Here we have written p, the *dipole moment* for qa. Thus at places far away from a dipole relative to the separation a, the field of relatively large charges close together is the same as for smaller charges at larger separation. All that matters is the product: charge $\times$ separation. This is the reason for using the dipole moment in discussing dipoles. Instead of the $1/r^2$ dependence of the field due to an isolated point charge, the field of a dipole falls off as $1/r^3$.

***Fig. 2.6** Torque on an electric dipole in a uniform electric field. Note positive direction of vector dipole moment* $\mathbf{p} = q\mathbf{a}$.



A very important property of an electric dipole is the torque exerted on it by a uniform electric field. Figure 2.6 shows a dipole in a constant field E. The torque τ is evidently given by

$$\tau = qaE \sin \theta$$

where θ is the angle between the dipole axis and the field. Again we may replace qa by p, to give

$$\tau = pE \sin \theta \tag{2.6a}$$

The dipole moment may be treated as a vector quantity $\mathbf{p} = q\mathbf{a}$ directed from the negative to the positive charge. We see that the torque on the dipole due to the external electric field tends to align the dipole moment parallel to the electric field.

d Field due to a plane distribution of charges Suppose we have a plane area on which charges are distributed uniformly with a surface density σ coulombs/m^2. We wish to calculate the field at a point P a distance a from the plane as shown in Fig. 2.7. We assume that the dimensions of the plane are much greater than a. We use Eq. (2.2) to add up the vector contributions of all charges at the point P. We begin by calculating the contribution from the ring of charge of radius r and width dr and then integrate for all such rings that make up the total plane charge distribution. The calculation of the contribution of charges on the ring is really a two-dimensional integration problem, but because of the symmetry we can reduce it to a simple summation. The contribution of an ele-

ment of charge on the ring to the field at P makes an angle θ with the x axis as shown on the figure. Because of symmetry, however, components of $\mathbf{E}$ perpendicular to the x axis cancel. Thus we need

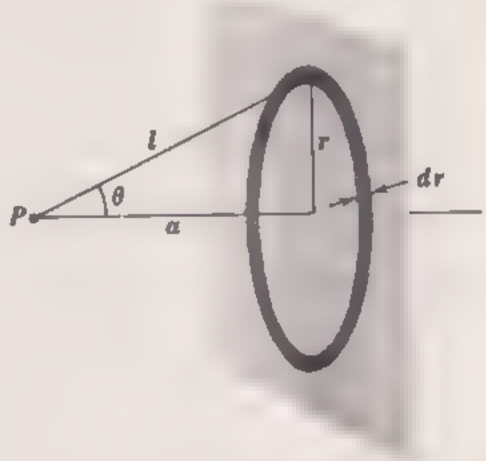


Fig. 2.7 The electric field at P from a plane distribution of charge is obtained by integrating the contributions from concentric rings.

consider only the x components of $\mathbf{E}$ at P. We can then use Eq. (2.2) to write, for the field at P due to the ring of charge,

$$dE = \frac{1}{4\pi\epsilon_0}\frac{\sigma 2\pi r\,dr}{l^2}\cos\theta$$

In order to reduce this to a single variable, we substitute $a\tan\theta = r$ and $a/\cos\theta = a\sec\theta = l$. By differentiation we have $dr = a\sec^2\theta\,d\theta$. Making these substitutions, we get

$$dE = \frac{\sigma}{2\epsilon_0}\frac{\tan\theta\sec^2\theta\cos\theta}{\sec^2\theta}\,d\theta = \frac{\sigma}{2\epsilon_0}\sin\theta\,d\theta$$

This is the contribution to the field at P from the ring we have chosen. To obtain the total field at P from all rings making up the plane charge distribution, we integrate this expression over the entire plane. The limits of integration are from $\theta = 0$ to $\theta = \pi/2$. Thus we find

$$E = \frac{\sigma}{2\epsilon_0}\int_0^{\pi/2}\sin\theta\,d\theta = -\frac{\sigma}{2\epsilon_0}[\cos\theta]_0^{\pi/2} = \frac{\sigma}{2\epsilon_0}$$

The resultant field is in the x direction and is independent of the distance from the plane as long as the plane is very large compared with a.

2.4 Conductors and Electric Fields

We now consider the electric field in the vicinity of a conducting body. The field *inside* a piece of metal is zero in the equilibrium state; otherwise, the freely moving charges, which account for its conducting properties, would move under the influence of the field until they were so arranged as to cancel the field. Thus, the only equilibrium condition for conduction electrons is one in which the average resultant field due to both the external charges and the redistributed charges in the metal is zero.

It follows from this that the direction of the electric field outside a metal body is everywhere normal to the metal surface. If this were not so, there would be a component of field along the surface. Such a field component along the surface would again cause motion of conduction electrons until the field component would be reduced to zero by the new distribution.

Finally, we can now argue that electric fields cannot penetrate conductors (since they must go to zero inside). Conductors thus make excellent shields for electric fields.

Although the conclusions thus reached are satisfactory from a practical point of view, two factors have been neglected. The first is that we are considering the static situation. With a rapidly varying field the situation is more involved. Second, the implied model of a conductor here is a grainless volume distribution of positive charge sharing its space with an equal volume distribution of movable negative charge. This is satisfactory for present purposes, even though far from realistic, but it is actually just the graininess (due to its atomic nature) that accounts for many interesting properties of matter.

2.5 Lines of Force

For problems involving fields, the idea of *lines of force* is a very useful approach, which depends intimately on the fact that the force between point charges depends on the inverse square of the distance of separation. In fact, there is a complete equivalence between the lines-of-force concept and the inverse-square law. Given the idea of lines of force as defined below, we can show with no additional assumptions that the inverse-square law follows. Similarly, this law is experimental justification for the use of lines of force.

We first describe the idea of lines of force, then show some

simple examples of the kind of reasoning they allow, and, finally, discuss their connection with the inverse-square law.

Lines of force are lines that originate only on positive electric charges and end only on negative charges. Lines are thus continuous, except at their *sources* and *sinks*, on positive and negative charges, respectively. The number of lines originating and terminating on charges is proportional to the magnitude of the charges. The *direction* of these lines is everywhere the same as the direction of the electric field, and the *density* of lines in a given region is a measure of the *magnitude* of the electric field. The density of lines means the number of lines per unit area threading a surface perpendicular to the direction of the lines at any given

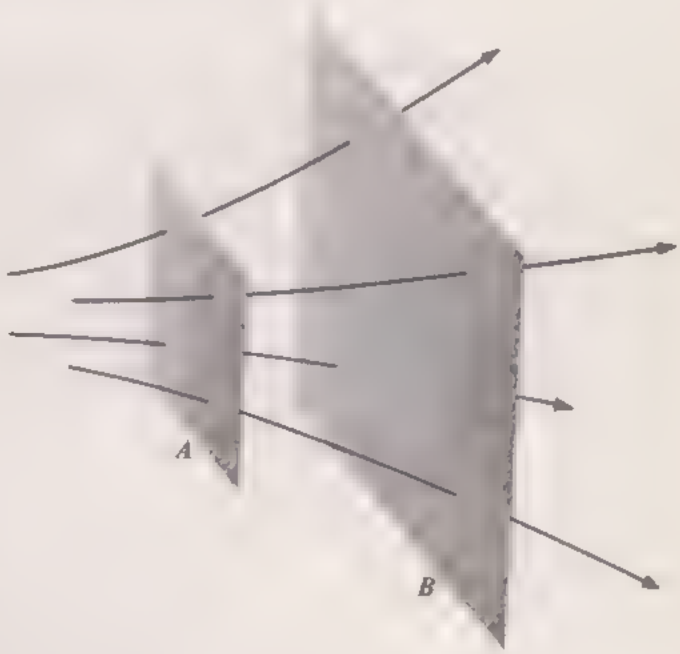


***Fig. 2.8* Lines of force cutting surface areas *A* and *B*. The density of lines and hence the field is greater at *A* than at *B*.**

point. Thus in Fig. 2.8 the density of lines of force through plane A is greater than the density at plane B, so the magnitude of the field at A is greater than at B. A sketch of some of the lines of force between a positive and a negative point charge is shown in Fig. 2.9.

This diagram provides another way of describing certain types of vector fields. Its use requires that the lines of force be continuous. It is quite different from our earlier picture (Fig. 2.2) used to describe a vector field.

In order to use lines of force in a quantitative way, we must assign a value to the number of lines originating per unit charge. This then defines the relationship between line density and electric

field magnitude. Let us clarify this by discussing the lines originating on an isolated point charge q_0 (coulombs). Some of these lines are sketched in Fig. 2.10.

We define unit field (1 newton/coulomb) arbitrarily as corresponding to unit density of lines of force (one line per square

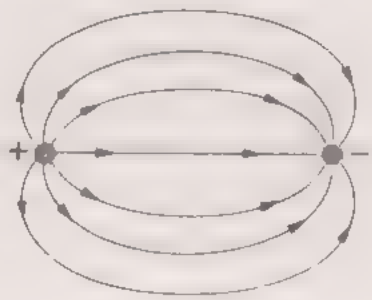


***Fig. 2.9** Some lines of force around a dipole.*

meter). We can then determine the required number of lines per unit charge, using the sketch and some quite simple reasoning. We draw (or imagine) a spherical surface of unit radius (1 m) around the charge as its center. From the symmetry of the geometric situation for an isolated point charge or from application of Eq (1.5) we know that the field is directed everywhere radially outward from the point charge. This must mean that the lines of force here are radial straight lines. The density of these lines where they

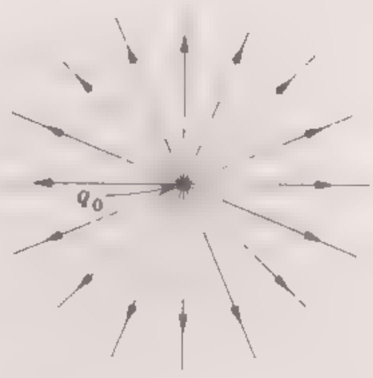


***Fig. 2.10** Lines of force around an isolated point charge.*

cut the spherical surface is uniform (by symmetry). The electric field E at the surface is calculated by $E = \dfrac{1}{4\pi\epsilon_0}\dfrac{q_0}{l^2}$, and the total area of the sphere is 4π. Thus the field at the unit sphere surface is $q_0/4\pi\epsilon_0$, which is also the density of lines of force according to our definition. Multiplying this by the area of the sphere, we find the total number of lines originating at q_0 and cutting the spherical

surface $= q_0/\epsilon_0$ Thus our unit field definition leads to $1/\epsilon_0$ lines of force per unit charge.

One use of lines of force is to give a qualitative picture of the electric field distribution. Figure 2.8 shows a two-dimensional plot of the lines of force between two equal and opposite point charges. Solving for the magnitude and direction of the field at a few points and knowing that lines of force are continuous make it possible to sketch such a set of lines. Such a diagram not only gives accurate information as to the field direction but also gives qualitative information about relative magnitude of the *field strength* at various points.

Finally, for certain situations of fairly high symmetry, the concept of lines of force allows very simple solution of certain field problems. The mathematical formalism of this method is known as Gauss' law and is discussed below.

There is a simple connection between the inverse-square law and lines of force. Again consider the case of an isolated point charge as shown in Fig. 2.10. Let us calculate the field at a distance of 2 m, using the lines-of-force point of view. We have seen already that the density of lines at 1 m is $q_0/4\pi\epsilon_0$. As shown in Fig. 2.11, the area of spherical surface subtended by a bundle of lines

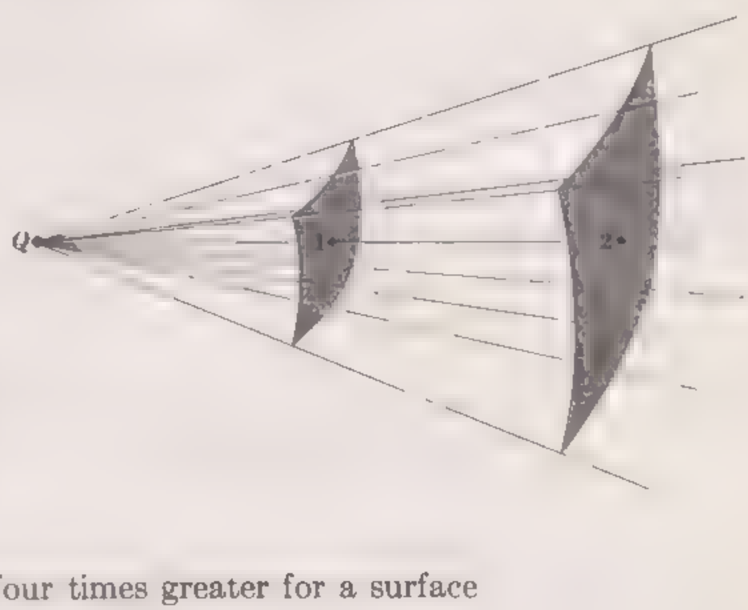


***Fig. 2.11** Elements of spherical surfaces intersected by a bundle of lines of force originating at a point charge.*

originating on the point charge is four times greater for a surface 2 m away than for a surface at 1 m distance. Therefore the density of lines of force at 2 m is just one-quarter that at 1 m, in exact agreement with the result if calculated from the field equation. Thus, at least for an isolated point charge, the field at any point in its vicinity can be calculated equally well from the density of

lines of force or from the inverse-square law. This result can be connected with the behavior of solid angles, as discussed in Sec. 2.6.

2.6 Gauss' Law

The treatment of more complicated situations not involving spherical symmetry is best done using Gauss' law. We should stress, however, that this actually embodies nothing more than the idea of the validity of the lines-of-force point of view. The theorem is stated mathematically as follows:

$$\iint_{CS} E_n \, dS = \sum_i \frac{q_i}{\epsilon_0}$$

or more simply, using vector notation,[1]

$$\iint_{CS} \mathbf{E} \cdot d\mathbf{S} = \sum_i \frac{q_i}{\epsilon_0} \tag{2.7}$$

Here the scalar quantity dS is the element of area. The vector $d\mathbf{S}$ is the vector element of area, having the same magnitude as dS and pointing in the outward normal direction from the surface. The subscript $\int_{CS}$ means that the integration is over a closed surface containing all the charges indicated by the summation $\sum_i q_i/\epsilon_0$.

Surface integral, scalar or dot product, and solid angle The *surface integral* $\int \mathbf{E} \cdot d\mathbf{S}$ is called the *flux* through the surface over which the integral is calculated. Figure 2.12 is a sketch of a charge q_0 and the solid angle that contains some of the lines of force from it that pass through a surface element

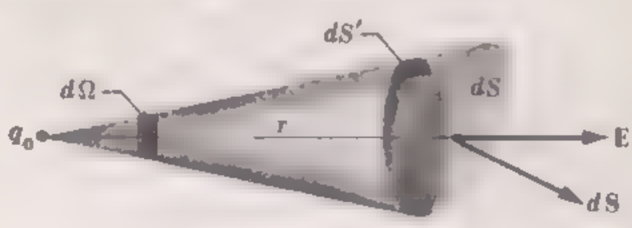


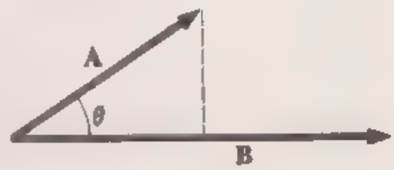
***Fig. 2.12** Graphical explanation of $\mathbf{E} \cdot d\mathbf{S}$, an element of the surface integral.*

[1] In order to simplify the writing of integrals, surface integrals such as this one will be written with only one integral sign. It will be understood that when the integral is taken over an area $d\mathbf{S}$, the double integral is implied. Similarly, the volume integral $\iiint dV$ will be written as $\int dV$, with the triple integral implied.

dS We shall now argue that the element of the surface integral $\mathbf{E} \cdot d\mathbf{S}$ is simply the number of lines of force going through the element of area dS.

As a first step we define the meaning of the *scalar* or *dot product* of two vectors. This is the product of one vector magnitude and the magnitude of the parallel component of the other vector. Thus in Fig. 2.13, $\mathbf{A} \cdot \mathbf{B} =$

Fig. 2.13 *The scalar or dot product of two vectors* **A** *and* **B** *is given by* $\mathbf{A} \cdot \mathbf{B} = AB \cos \theta$.



$AB \cos \theta$, where $A \cos \theta$ is the component of A parallel to B; alternatively, $B \cos \theta$ is the component of B parallel to A. The scalar product is not itself a vector, so these two descriptions are equivalent.

We now apply this scalar product to the case at hand, using Fig. 2.12. Reverting to the vector notation for the electric field (rather than using lines of force), we draw the vector $\mathbf{E}$ in line with the radius vector from q_0. Similarly, we draw the normal vector $d\mathbf{S}$ to the surface element dS as a vector with a magnitude representing the area of the element. The scalar product $\mathbf{E} \cdot d\mathbf{S}$ then means $E\, dS \cos \theta$ or $E_n\, dS$, where E_n is the component $E \cos \theta$ of $\mathbf{E}$ parallel to $d\mathbf{S}$, or normal to the area dS. The alternative description is $E\, dS'$, where dS' is the projection $dS \cos \theta$ of the area, perpendicular to $\mathbf{E}$. Now if we use this last description and go back to the lines-of-force description, we may interpret $\mathbf{E} \cdot d\mathbf{S}$. Thus, since the magnitude of $\mathbf{E}$ is measured by the density of lines, $\mathbf{E} \cdot d\mathbf{S}$ is just the total number of lines going through the area dS. This is the basic meaning of the surface integral.

We may also use Fig. 2.12 to discuss the significance of a solid angle. The surface element dS subtends a solid angle $d\Omega$ at the charge q_0. A solid angle is the analogue in three dimensions of the usual angle measurement in two dimensions. In two dimensions we speak of the angle subtended at the center of a circle by a given arc of the circle. Unit angle, the radian, is the angle subtended by an arc equal in length to the radius. In three dimensions, a given area on a sphere subtends a certain solid angle at the center of the sphere. This solid angle is independent of the shape of the area and depends only on the magnitude of the area. Unit solid angle is that subtended at the center of a sphere by an area r^2, on a sphere of radius r. The unit of solid angle is called the *steradian*. The total solid angle is 4π steradians (since the total area of a sphere is $4\pi r^2$); dS and dS' in Fig. 2.12 subtend the same solid angle. The idea of a solid angle can be applied generally, not merely when studying a sphere. Thus any surface dS subtends a solid angle $d\Omega$ at a point a distance r away, where

$$d\Omega = \frac{dS \cos \theta}{r^2}$$

Here θ is the angle between the radius vector r from the point and the normal to the surface, so $dS \cos\theta = dS'$ is the projection of the surface element perpendicular to the line from the point in question.

With the foregoing mathematical interpretation we can immediately understand the meaning of Gauss' law as we have written it in Eq. (2.7). Thus the integration of $\mathbf{E}\cdot d\mathbf{S}$ over an entire closed surface gives the flux or the net number of lines emerging through the surface. (Lines pointing inward through the surface make a negative contribution to the flux since θ is the angle between the *outward* drawn normal and the field direction.) Gauss' law says that the *net* number of lines emerging through a closed surface depends only on the total charge surrounded by that surface (with the proportionality factor $1/\epsilon_0$). Thus in Eq. (2.7), $\sum_i q_i/\epsilon_0$ refers to all the charges q_i *inside* the volume surrounded by the closed surface over which the surface integral is performed.

The importance of Gauss' law is related to the fact that the surface integral of $\mathbf{E}$ over an enclosing surface is equal to the number of lines of force emerging for *any* shape of surface whatever, as we now show. Consider a charge q_0 entirely surrounded by any shape of closed surface (Fig. 2.14). Consider the element of surface area subtended by a small solid angle $d\Omega$ at the charge position. Let r be the distance from charge to surface element. Using Fig. 2.12 as an enlarged view, we write the surface element dS' in terms of the solid angle,

$$dS' = r^2\, d\Omega = dS \cos\theta$$

Now E_n, the normal component of the electric field, is $E\cos\theta$. We may thus write Eq. (2.7) as

$$\int_{CS} E\cos\theta\, dS = \frac{q_0}{\epsilon_0}$$

E at the surface dS due to q_0 is given by $E = \dfrac{1}{4\pi\epsilon_0}\dfrac{q_0}{r^2}$, and as seen above, $dS = r^2\, d\Omega/\cos\theta$. Combining these facts, we find

$$\int_{CS} E\cos\theta\, dS = \frac{1}{4\pi\epsilon_0}\int_{CS}\frac{q_0}{r^2}\, r^2\, d\Omega = \frac{q_0}{4\pi\epsilon_0}\int_{CS} d\Omega$$

But $\int d\Omega$ over the entire surface is 4π, whatever its shape, so long

as it is closed, so we have shown that Gauss' law is valid for a single point charge. It is also true for any amount of charge distributed in any way inside the surface. This follows if we again relate Gauss'

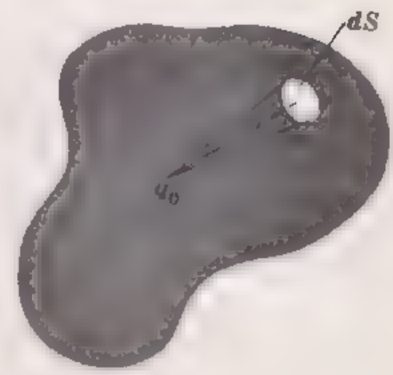


***Fig. 2.14** An element dS of surface subtended by a small solid angle at the charge q_0.*

law to the net number of lines of force coming out of the enclosing surface. Since this number depends only on the size of each charge q and not on its position inside the surface, the net flux of lines outward must be equal to $\Sigma q/\epsilon_0$.

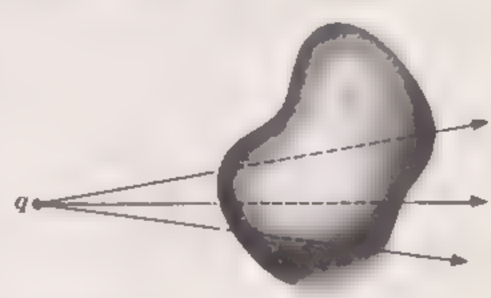
An alternative and sometimes more useful form of Gauss' law is given in terms of the charge density ρ inside the closed surface.

$$\int_{CS} \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0}\int_v \rho\, dV \tag{2.8}$$

where dV is an element of volume. The volume integral on the right-hand side gives the total charge inside the volume enclosed by the closed surface, so the meaning here is the same as before.

The physical ideas involved in Gauss' law are the same as those involved when the ideas of lines of force are used. Take, for ex-

***Fig. 2.15** The net number of lines emerging from a volume due to a charge outside the volume is zero. (The number of lines entering must just equal the number emerging.)*



ample, a closed surface which surrounds no charge but which has charges outside it. From either approach we see that the net number of lines (the number going out less the number going in) coming out through the surface is zero (see Fig. 2.15).

The usefulness of the concept of lines of force as an aid to

thinking does not imply their reality as physical entities. They are merely a graphical way of picturing the mathematical facts resulting from the inverse-square law.

2.7 Examples, Use of Gauss' Law

In a large class of problems the electric field can be found directly from Gauss' law. In these cases surfaces can be found over which the field is constant in magnitude and makes a constant angle with the surface. We give a number of examples.

a Field of a point charge This is quite simple, but it is helpful to illustrate the approach used later in more complicated situations. As shown in Fig. 2.16, a sphere is drawn around the point charge Q

Fig. 2.16 Sphere drawn around point Q to allow calculation of the field at a distance r, using Gauss' law.

at a radius r. By symmetry we know that the field is the same at every point on this surface (and directed outward if Q is positive). Using Gauss' law,

$$\int_{CS} \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0} Q$$

but since $\mathbf{E}$ is uniform over the surface and everywhere normal to the surface, so that $\cos\theta = 1$, we can write

$$E = \frac{(1/\epsilon_0)Q}{\int_{CS} dS} = \frac{Q}{4\pi\epsilon_0 r^2}$$

Another fact that is now obvious is that the field configuration *outside* any spherical distribution of charge is exactly equivalent to the field that would result if all the charge were concentrated at the center. This may be shown either from the result of Gauss' law or by thinking in terms of lines of force. As shown in Fig. 2.17, the total number of lines is the same in both cases, and since they both

emerge radially and uniformly, the line configuration and hence the field outside must be identical. This reasoning works for any spherically symmetric distribution, which requires that the charge density be a function of the radius alone.

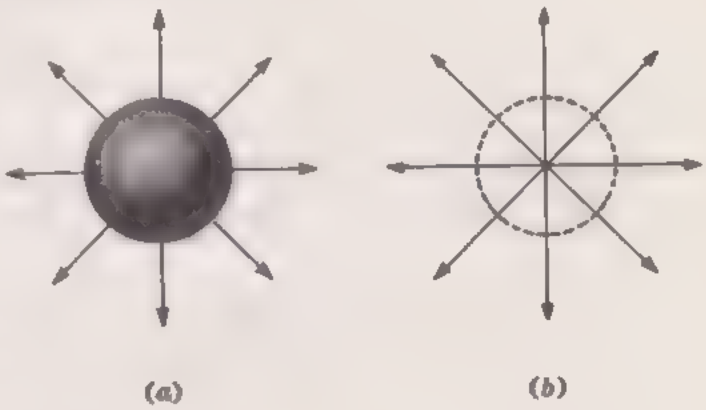


***Fig. 2.17** (a) Lines of force outside a spherical charge distribution; (b) lines of force from a point charge.*

b Field around a long uniform linear charge distribution This is the same problem considered earlier, which we now solve using Gauss' law. Let the charge density be μ coulombs/m. Draw a cylinder of radius r and length l around the charge, as shown in Fig. 2.18. By

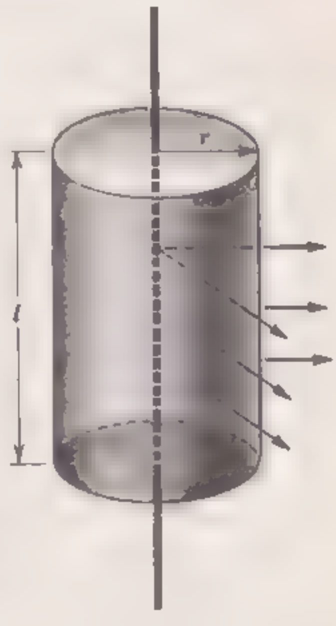


***Fig. 2.18** Field around a linear charge distribution, showing a few of the radial lines of force. The cylindrical Gaussian surface has a radius r.*

symmetry, all the lines of force go radially outward. Thus no lines cross the flat ends of the cylinder, and the integral $\int \mathbf{E} \cdot d\mathbf{S}$ over these ends is zero. Over the remaining (curved) surface of the cylinder the field is uniform and outward (perpendicular to the

surface). We can apply Gauss' law formally then by writing

$$\int_{CS} \mathbf{E}\cdot d\mathbf{S} = \int_{\text{flat surfaces}} \mathbf{E}\cdot d\mathbf{S} + \int_{\text{curved surface}} \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0}\mu l$$

Since the flat-surface term contributes nothing, we may write

$$\int_{CS} \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0}\mu l$$

where the integration is over the curved surface only. Since E is uniform over the curved surface, we may write

$$E = \frac{1}{\epsilon_0}\frac{\mu l}{\int dS} = \frac{\mu l}{2\pi r l \epsilon_0} = \frac{\mu}{2\pi\epsilon_0 r}$$

as found before.

c *Field out from an infinite plane of uniform charge density* Let the surface charge density be σ coulombs/m^2. Draw a pillbox-shaped surface of cross section A as in Fig. 2.19. The charge contained inside will be σA, and by symmetry there will be no flux coming out of the curved surface (i.e., $\int \mathbf{E}\cdot d\mathbf{S} = 0$ over the curved surface). Only the two flat ends contribute, so we find

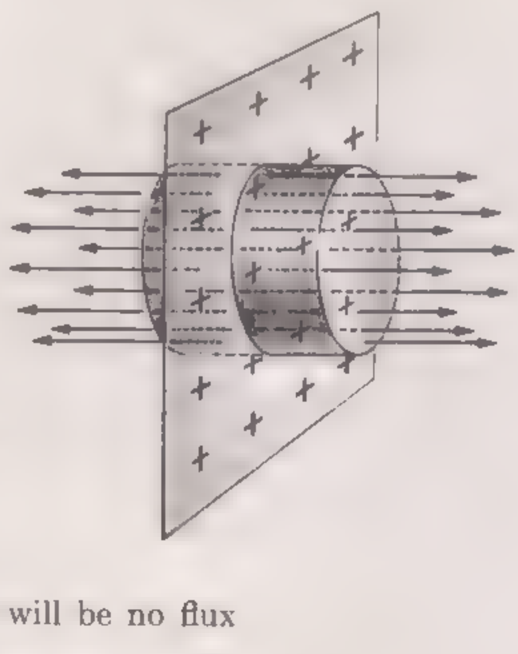


Fig. 2.19 Calculation of field from uniform plane distribution of charge. Pillbox-shaped Gaussian surface.

$$E = \frac{1}{\epsilon_0}\frac{\sigma A}{2A} = \frac{\sigma}{2\epsilon_0}$$

This result is identical with our earlier calculation of Example 2.3*d* in which the field was obtained by direct integration of contributions from each charge element. Gauss' law often provides a short cut to such solutions in situations of high symmetry.

Let us examine in some detail the alternative use of direct integration and Gauss' law in some situations involving plane distributions of charge in the presence of conductors. First take the case shown in Fig. 2.20*a* of a plane conductor with a surface

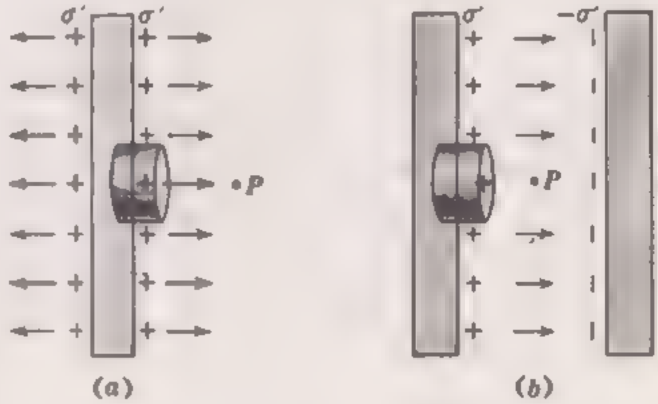


***Fig. 2.20** Electric field in the vicinity of plane distributions of charges on a conductor. (a) Plane conductor, uniformly charged. (b) Two parallel plane conductors, oppositely charged.*

charge density σ' coulombs/m^2 on each side. We first calculate the field at P by means of the vector sum of the contributions from each charge. Using the result of Example 2.3*d* for the contribution from a plane of charges, we find that the contribution at P of the surface charge on the right-hand surface is $E_1 = \sigma'/2\epsilon_0$. Similarly, the left-hand surface contributes another field in the same direction, $E_2 = \sigma'/2\epsilon_0$. The sum of these is $E = E_1 + E_2 = \sigma'/\epsilon_0$. Note that we have considered *all* charges in obtaining this result.

We now arrive at the same result by using Gauss' law. We draw the Gaussian surface as shown; however, the lines of force emerge from only the right-hand flat surface of the Gaussian pillbox since there can be no field and hence no lines of force within the metal plate. Use of Gauss' law then gives

$$E = \frac{1}{\epsilon_0}\frac{\sigma' A}{A} = \frac{\sigma'}{\epsilon_0}$$

in agreement with the detailed calculation. In this calculation the charges on the left-hand surface of the metal are not explicitly considered. However, it is the presence of the charges on the left-hand side of the metal plate that allows the field inside the plate to be zero. That is, the field inside the metal is the sum of the two

equal and opposite uniform fields of the charges on the two surfaces and is hence equal to zero.

Now consider the pair of conducting plates shown in Fig. 2.20*b*, where we assume that the separation between the plates is much less than their dimensions, so they look like infinite planes from the point P between plates. In this situation, the attractive forces between opposite charges draw all the excess charges to the inner surfaces as shown. We assume equal charge densities $\pm\sigma'$. We first calculate the field at P by means of the detailed summation method of Eq. (2.2). From our previous result, the contribution from the left-hand positive charges is $E_1 = \sigma'/2\epsilon_0$. The negative charges on the right-hand plate give an equal field, $E_2 = \sigma'/2\epsilon_0$, in the same direction. The total field is then $E = E_1 + E_2 = \sigma'/\epsilon_0$.

In order to solve the same problem by Gauss' law, we draw a Gaussian surface at, say, the left-hand plate and, since the lines of E are parallel to each other and perpendicular to the surface, find for the field at P,

$$E = \frac{1}{\epsilon_0}\frac{\sigma' A}{A} = \frac{\sigma'}{\epsilon_0}$$

In this formulation, the negative charge distribution on the right-hand plate has not been considered explicitly, although of course its presence accounts for the collection of positive charges on one side of the conductor only. The same result would of course be found if Gauss' law had been applied to the negative charges only. Again, the zero field inside the plates is accounted for by the cancellation of equal and opposite uniform fields due to the two plane charge distributions.

The general result is that the uniform field in the region near a uniform-plane charge distribution is $E = \sigma/2\epsilon_0$, but when the charges are at the surface of a conductor, the field is given by $E = \sigma/\epsilon_0$; the factor of 2 results from the absence of field within the conductor.

d Field inside a hollow conductor The experimental accuracy for which the inverse-square law is known is remarkably great. This very satisfactory situation results from experiments done inside hollow conductors. It turns out that the field inside a hollow conductor is zero if, and only if, the law of force is accurately inverse-square.

We may take the special case of a hollow conducting sphere.

The field at any point C inside this sphere (Fig. 2.21), due to a charge density σ on the surface, is to be shown equal to zero. We must not invoke the concept of lines of force or any other consequence of the inverse-square law, since that is what we wish to

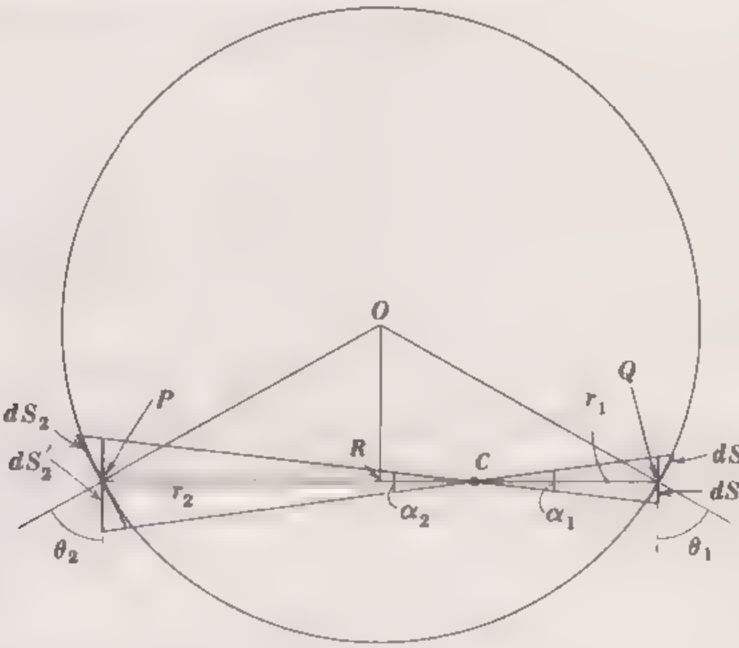


***Fig. 2.21** Construction showing absence of electric field inside a hollow spherical conductor.*

prove. We may, however, assume uniform charge distribution over the sphere, by using the argument of symmetry. The field at C is obtained by taking the vector sum of all the contributions at the point C. We may start by considering the contributions from the charges $\sigma\, dS_1$ and $\sigma\, dS_2$ subtended by the equal and opposite infinitesimal solid angles α_1 and α_2. From the properties of solid angles we know that

$$\frac{dS_2 \cos \theta_2}{r_2^{\,2}} = \frac{dS_1 \cos \theta_1}{r_1^{\,2}}$$

Furthermore, for small solid angles $\theta_1 = \theta_2$.[1] Thus,

$$\frac{dS_2}{r_2^{\,2}} = \frac{dS_1}{r_1^{\,2}}$$

or

$$\frac{\sigma\, dS_2}{r_2^{\,2}} = \frac{\sigma\, dS_1}{r_1^{\,2}}$$

[1] This may be shown as follows: dS_1' and dS_2' are $\perp PQ$.
Draw $OR \perp PQ$.
OR is also the bisector of $\angle POQ$.
$\therefore \theta_2 = \angle ROP$, $\theta_1 = \angle QOR$.
$\therefore \theta_1 = \theta_2$.

If the inverse-square law holds, the components of the field at C are just

$$dE_1 = \frac{1}{4\pi\epsilon_0}\frac{\sigma\, dS_1}{r_1^{\,2}} \qquad \text{and} \qquad dE_2 = \frac{1}{4\pi\epsilon_0}\frac{\sigma\, dS_2}{r_2^{\,2}}$$

and these are in opposite directions. We see thus that for the particular part of the total charge density subtended by the pair of equal and opposite elements of solid angles, the contributions to the field just cancel to give a net value of zero. Since the remaining surface of the sphere can be similarly broken up into areas subtended by pairs of equal and opposite solid angles, it follows that the total contribution to the field at C is zero, if, that is, the inverse-square law is correct.

The experimental proof of this, as carried out by Maxwell, involved placing a concentric conducting sphere inside another one, and insulated from it. Starting with no net charge on either sphere, a large charge was then placed on the outer sphere, and a field was looked for in the space between the two spheres. As we shall see in the next chapter, if a radial electric field were present between the spheres, a *potential difference* would be set up between them. Only if *no* field is produced inside the outer sphere is the potential difference between the spheres zero. No potential difference was found, and Maxwell was able to state that in his experiments the power 2 in the inverse-square law was accurate to at least 1 part in 21,600. Modern experiments involving the same principles have increased the accuracy of this to 1 part in 10^9.

PROBLEMS

2.1 Eight identical point charges of Q coulombs each are placed at the corners of a cube whose sides have a length of 10 cm.

- *a* Find the electric field at the center of the cube.
- *b* Find the electric field at the center of a face of the cube.
- *c* Find the field at the center of the cube if one of the corner charges is removed.

2.2 A charge of Q coulombs is at the center of a sphere of radius 2 m.

- *a* How many lines of force originate on the charge?
- *b* How many lines of force emerge through an area of ½ m^2 of the surface of the sphere?
- *c* What is the density of lines of force for unit electric field?
- *d* What is the field at the surface of the sphere?

2.3 Five thousand lines of force enter a certain volume of space and three thousand lines emerge from it. What is the total charge in coulombs within the volume?

2.4 Lines of force emerge radially from a spherical surface and have a uniform density over the surface. What are the possible distributions of charge within the sphere?

2.5 A thin circular ring of radius 20 cm is charged with a uniform charge density of μ coulombs/m. A small section of 1 cm length is removed from the ring. Find the electric field intensity at the center of the ring.

2.6 A circular disk of 10 cm radius is charged uniformly with a total charge of Q coulombs. Find the electric field intensity at a point 20 cm away from the disk, along its axis.

2.7 A total charge of Q coulombs is uniformly distributed over the volume of a sphere of 20 cm radius. Find the electric field intensity:

a At the center of the sphere
b At a point 10 cm from the center of the sphere
c At a point on the surface of the sphere
d At a point 50 cm from the center of the sphere

2.8 A total charge of Q coulombs is uniformly distributed along a rod 40 cm in length. Find the electric field intensity 20 cm away from the rod along its perpendicular bisector, as shown in Fig. P2.8.

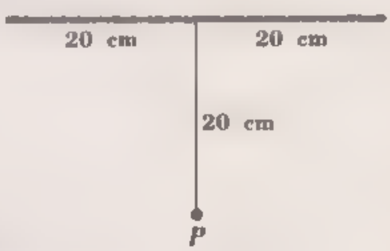


Fig. P2.8

2.9 A semicircular rod as shown in Fig. P2.9 is charged uniformly with a total charge of Q coulombs. Find the electric field intensity at the center of curvature.

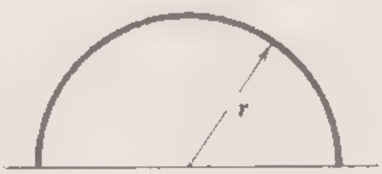


Fig. P2.9

2.10 Two identical point charges of $+Q$ coulombs are separated by a distance of 10 cm, as shown in Fig. P2.10. Calculate the work per unit charge to bring up another charge from far away along the perpendicular bisector of the line joining the two charges to the point midway between the two charges.

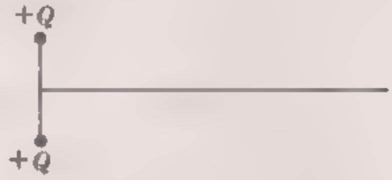


Fig. P2.10

2.11 If one of the two charges in Prob. 2.10 is changed from $+Q$ to $-Q$, calculate the work per unit charge to bring up a charge to the same position as in Prob. 2.10.

2.12 A dipole having a dipole moment $p = Qa$ coulomb-m makes an angle θ with the direction of a uniform electric field E, as shown in Fig. P2.12.

a Calculate the torque on the dipole.

b Find the work necessary to reverse the position of the dipole from its equilibrium position along the field to the opposite direction.

c For small amplitudes of oscillation about its equilibrium position, calculate the period of oscillation of the dipole if it has a moment of inertia I about its center.

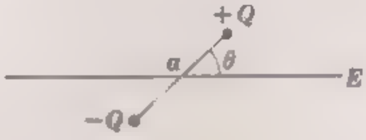


Fig. P2.12

2.13 A dipole of moment $p = Qa$ coulomb-m is aligned parallel to an electric field along the x axis. The field is nonuniform and varies in magnitude linearly along the x axis with a slope $dE/dx = K$. Find the force on the dipole.

2.14 A pair of electric dipoles such as shown in Fig. P2.14 is called a *quadrupole*. Find the electric field at a point P along the axis of the quadrupole at a distance r ($r \gg a$) from its center.

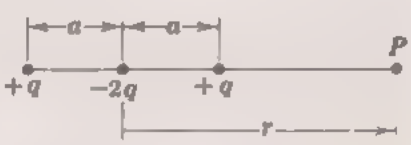


Fig. P2.14

2.15 A thin hemispherical cup (an insulator) of radius R bears a charge Q uniformly distributed over its surface. Find the electric field at the center of the flat surface of the hemisphere.

2.16 Consider a charge Q distributed through a sphere of radius R with a density

$$\rho = A(R - r) \qquad 0 < r < R$$

when ρ is in coulombs per cubic meter. Determine the constant A in terms of Q and R. Calculate the electric field inside and outside the sphere.

THREE

The Electric Potential

3.1 Introduction

Now that we have discussed Coulomb's law of force between charges in terms of the electric field, we are ready to take another step that simplifies many considerations. As in the study of mechanics, we often find it useful to think in terms of the *work* done by electric forces, and we further find that very often the ideas of *potential energy* are powerful aids to understanding the behavior of electric charges. This chapter is devoted to building up the necessary new ideas and to illustrating some of the ways of using them.

3.2 The Line Integral, Work

In mechanics, the *line integral* plays a very fundamental role in problems involving work. Its importance is equally great in electrostatics, so we discuss it briefly here. Figure 3.1 recalls the simple idea that the work done on a body depends on the component of

force acting in the direction of motion times the distance traveled. In differential form, the element of work is

$$dW = F \cos\theta \, dl$$

where dl is the element of distance along the direction of motion

Fig. 3.1 Use of line integral for the calculation of work. Element of work $dW = F \cos\theta \, dl$.

and F is the force acting on the body. For a finite amount of motion, the calculation is

$$W = \int_A^B (F \cos\theta) \, dl$$

where we integrate from the beginning to the end of the path from A to B. In vector notation, this is

$$W = \int_A^B \mathbf{F} \cdot d\mathbf{l} \qquad (3.1)$$

This is another use of the scalar product of two vectors, discussed in Sec. 2.6, where it was applied to the surface integral. However, the situation here is physically very different. The integral for flux is taken over a surface and the integral for work is taken along a line, so there is no possibility of confusion.

The line integral and the ideas of work and energy are useful when we have a *conservative* system. As an illustration we may consider the earth's gravitational field, a vector field for which there is a force per unit mass at each point in space. This is called a conservative field because in it the energy associated with position is conserved. Thus the work we do in lifting a body is independent of the path taken, and, furthermore, we can get back all the energy thus expended by letting the body return to its starting point. The energy stored by virtue of position is the potential energy.

The requirement for a conservative system is that the potential energy of a body in the field be defined uniquely by its position. This is true if the external work necessary to move a body from one

point to another is independent of the path taken between the two points. Only under this condition is the work to bring a body to a particular position from a reference point unique, and therefore only in this case is the potential energy uniquely defined. We shall show that any possible space arrangement of electric field due to stationary charges is a conservative field. This is the reason that the use of work and energy is as natural in electrostatics as in mechanics.

In order to show that the inverse-square law of force of electrostatics gives rise to a conservative field, we begin with the field of a fixed charge Q at some point as shown in Fig. 3.2. We calculate

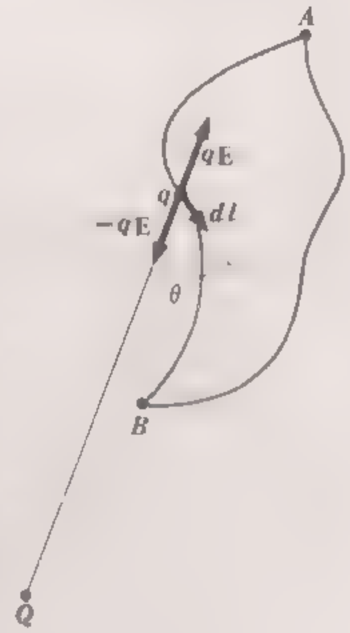


***Fig. 3.2** Line integral for calculation of external work done against the field of a point charge Q in moving charge from A to B. For a conservative field the work done by any path is the same. The external force is −qE.*

the *external* work that must be done on another charge q to move it *quasi-statically* from point A to point B. The quasi-static process is one in which the external force is only infinitesimally different in magnitude from the force of the field and is in a direction opposite to that of the field. Consequently, the motion of the body results from an infinitesimal net force, so that the body arrives at its new position B with no kinetic energy. This procedure allows us to avoid kinetic-energy considerations. We calculate the external work necessary to move the charge q along the differential distance dl and then integrate this expression over the entire path from A to B to find the total external work done. The force of the field of Q on the charge q is $q\mathbf{E}$, so the external force required is $\mathbf{F} = -q\mathbf{E}$. This force makes an angle θ with the path element dl. The work

done along dl is then $dW = F \cos\theta \, dl$, and the total external work from A to B is

$$W = \int_A^B F \cos\theta \, dl = -q \int_A^B E \cos\theta \, dl$$

From the figure we see that $\cos\theta \, dl = dr$, the change in the distance r between the two charges when we move q along dl. Making this substitution and replacing E by the law of force, we get

$$W = -\frac{qQ}{4\pi\epsilon_0} \int_A^B \frac{1}{r^2} \, dr = \frac{qQ}{4\pi\epsilon_0} \left(\frac{1}{r_B} - \frac{1}{r_A} \right)$$

This result shows that the work necessary to move q from A to B is independent of the particular path taken and depends only on the positions A and B. We have thus proved that the field of a point charge is conservative. It follows that any field configuration made up of the superposition of the fields of any arbitrary distribution of point charges will also give a conservative field. This is true because work is a scalar quantity, so that the total work done in moving the charge q from A to B against the fields of a distribution of charges is just the sum of individual terms such as we have calculated for the field of the point charge Q. Since each term is independent of the path taken, the sum of all terms is also independent of the path.

We state this result more formally by writing for the external work

$$W = -q \int_A^B \mathbf{E} \cdot d\mathbf{l}$$

Since q is constant, the condition we have proved for the electric field is that

$$\int_A^B \mathbf{E} \cdot d\mathbf{l} = \text{const} \tag{3.2}$$

This integral is independent of the path taken from A to B, for a static field arising from any charge distribution. An equivalent statement of the conservative nature of an electrostatic field is

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0 \tag{3.3}$$

that is, the line integral around a *closed* path equals zero. This follows at once since, if we take a charge q from A to B and then

return along an alternative path as shown in Fig. 3.2, the work done going from B to A must be just the negative of the work from A to B, so the total work around any closed path must be zero in a quasi-static process.

3.3 Potential Energy

When we have a conservative field, as we have just argued for a static electric field, we can speak usefully of the potential energy of a charge. Since the work to put it in a particular place is independent of the path taken, the potential energy is a function of position only. In mathematical terminology we may write for the difference in potential energy of a charge q at A and at B,

$$PE_B - PE_A = -\int_A^B q\mathbf{E}\cdot d\mathbf{l} \tag{3.4}$$

The work is given in joules if mks units are used.

This gives the external work to move the charge from A to B against the electric forces. Since this is a difference expression, only *differences* in potential energy can be calculated. Absolute values have no meaning.

One final remark calls attention to the fact that we have developed here a new kind of description of space. We have already seen that we can describe the electrical situation by means of a *vector electric field* $\mathbf{E}$, which gives magnitude and direction of the force per unit charge at each point in space. But now we are able to describe the same situation in terms of a *scalar* quantity, the amount of work necessary to place a given charge at any point in the space. We shall find important use for the relationship that must exist between these two kinds of description.

3.4 Potential Difference and Potential

It is convenient to discuss the difference in potential energy in terms of work on a *unit* positive charge. Thus,

$$\frac{PE_B - PE_A}{q} = -\int_A^B \mathbf{E}\cdot d\mathbf{l} \qquad \text{joules/coulomb} \tag{3.5}$$

This quantity is called the *potential difference* between points A and B, and its unit is the *volt*, in both the mks and practical systems of

units (see Chap. 15) Potential difference is often written as V_{AB} or $V_B - V_A$. Note that both **E** and V relate to effects on a unit charge. The electric field **E** gives the force on a unit charge, and the potential difference V_{AB} gives the external work necessary to move a unit charge from one position to another. To be rigorous, we can again use the differential form,

$$V_{AB} = \frac{d(PE_B - PE_A)}{dq}$$

to avoid any disturbing effects of a finite test charge on the original charge distribution.

Although it is only the *difference* in potential between two points that has fundamental significance, it is often convenient to choose (arbitrarily) a reference point for zero potential at infinity. When this is done, the *potential* V at a given point is defined as the external work necessary to bring a unit positive charge from infinity to the point in question. Thus the potential at a point P is given by

$$V_P = -\int_{\infty}^{P} \mathbf{E} \cdot d\mathbf{l} \tag{3.6}$$

The *potential* of a point is the potential difference between that point and a point at infinity.

Since potential difference is related to the line integral of **E**, there is a simple graphical relationship between V and **E**. We show this for the case of the field around a spherical shell of charge as in Fig. 3.3. In the graph below the sketch a plot of E against distance is

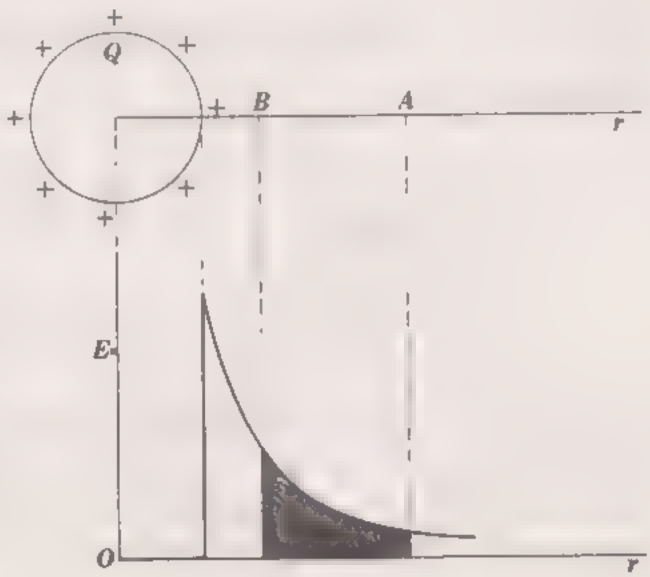


***Fig. 3.3** Plot of electric field of a uniform charge distribution on a spherical shell. Area under E curve between points A and B measures the potential difference between the two points.*

drawn. From Eq. (3.5) it follows that the potential difference between points A and B is simply the area under the curve of E versus r, between A and B. The potential of a point on the surface of the charged sphere is just the area under the E curve from infinity to the surface. In this simple example, **E** is along the path we take between A and B. In cases where this is not true, account must be taken of the angle between the field direction and the path, as required by the scalar product involved in the line integral. For example, if a straight-line path taken between the points A and B makes a constant angle θ with the field direction, the potential difference is given by the area under the curve of $E \cos \theta$ versus r between the two points.

3.5 *Examples, Calculation of Potential*

a Potential difference between two points in the region of a point charge Q (Fig. 3.4) We have

$$V_{AB} = V_B - V_A = -\int_A^B \mathbf{E} \cdot d\mathbf{r} \tag{3.5}$$

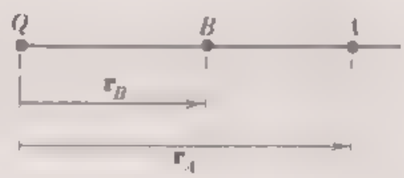


***Fig. 3.4** Calculation of the potential difference between points B and A due to the point charge Q.*

But **E** at a distance **r** from the point charge Q is by Coulomb's law $\frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}}$, and **E** is along **r** so that the $\cos \theta$ term implied by the dot product is 1, ∴

$$V_{AB} = -\frac{Q}{4\pi\epsilon_0}\int_A^B \frac{dr}{r^2}$$

$$= \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_B} - \frac{1}{r_A}\right) \qquad \text{joules/coulomb, or volts}$$

b Potential of a point P at a distance r from a point charge Q We mean by this the work per unit charge to bring a test charge up to P from infinity. We calculate this by using the result of the preceding problem. Let the point P be at B and let $A \to \infty$. This gives

$V_P = Q/4\pi\epsilon_0 r$ volts for the potential at a point P, r meters from a point charge. Thus the potential due to a point charge falls off as $1/r$. That the potential is spherically symmetrical around the point charge follows from the similar symmetry of the electric field. The result $V_P = Q/4\pi\epsilon_0 r$ is also valid for the potential outside any spherical charge distribution according to the reasoning given in Example 2.7*a*.

c Potential due to several point charges Suppose we calculate the potential at the origin due to charges q_1 at y_1 and q_2 at x_2 as shown

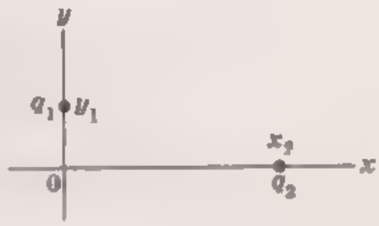


Fig. 3.5 ***Calculation of the potential at 0 due to several point charges. The potential is the scalar sum of terms due to individual charges.***

in Fig. 3.5. Since potential is a scalar quantity, we write directly, using the results of Example 3.5*b*,

$$V_0 = \frac{1}{4\pi\epsilon_0}\left(\frac{q_1}{y_1} + \frac{q_2}{x_2}\right) \qquad \text{volts}$$

Thus the general expression for the potential due to a distribution of point charges is

$$V = \frac{1}{4\pi\epsilon_0}\sum_i \frac{q_i}{r_i} \qquad \text{(scalar sum)} \tag{3.7}$$

If the distribution is continuous,

$$V = \frac{1}{4\pi\epsilon_0}\int_{\text{vol}} \frac{\rho\, dv}{r} \tag{3.8}$$

where ρ = density of charge, coulombs/m^3
dv = element of volume
r = distance from element to point in question

d Potential difference between two points out from an infinite uniform line of charge We proceed as before, using our earlier result

that the field is given by $E = \mu/2\pi\epsilon_0 r$ (Example 2.3*b*). Then if B is the point closest to the line of charges, we write

$$V_{AB} = -\int_A^B \mathbf{E}\cdot d\mathbf{r} = -\frac{\mu}{2\pi\epsilon_0}\int_A^B \frac{dr}{r} = \frac{\mu}{2\pi\epsilon_0}\ln\frac{r_A}{r_B}$$

Now if we attempt to calculate the potential (with respect to ∞) by letting r go to ∞ as in Example 2.3*b*, we find that V anywhere in the region of the linear distribution (r_B finite) goes to infinity. This is correct and results from the assumption of an infinite charge, as required to give a finite charge density μ over an infinite length. The same situation occurs for the potential out from an infinite plane with a uniform density, which would also require an infinite charge. Since in practical problems we are interested in potential *differences* between positions with finite separation and never in absolute potentials, this never causes trouble. Our choice of infinity for the point of zero potential was arbitrary, so any problem we solve by the use of potential will give the same answer regardless of our choice of reference for zero potential In the most general sense, potential is defined only within an arbitrary constant, so only potential differences have any real significance.

3.6 Potential Related to Electric Field

We now examine in more detail the relationship between electric field and potential given by the equation

$$V_{AB} = -\int_A^B \mathbf{E}\cdot d\mathbf{l} \qquad (3.5)$$

We first take the derivative of both sides to get

$$dV = -\mathbf{E}\cdot d\mathbf{l} \qquad \text{or} \qquad dV = -E\cos\theta\, dl$$

Thus,

$$\frac{dV}{dl} = -\mathbf{E}\cos\theta \qquad (3.9)$$

Now since $E\cos\theta$ is the component of $\mathbf{E}$ in the general direction dl, this equation tells that the negative of the rate of change of potential as we move in some direction dl is equal to the component of the electric field in that direction.

To be more explicit, suppose in a rectangular-coordinate system the field at the origin points in some general direction as shown

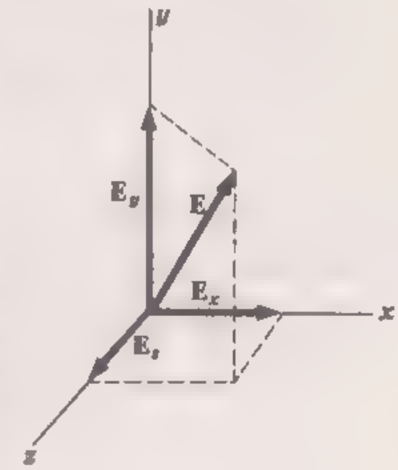


***Fig. 3.6** Components of the field can be obtained from the rate of change of the potential along given directions.*

in Fig. 3.6. Then the components of the field in directions x, y, and z are given by

$$-\frac{dV}{dx} = E_x$$

$$-\frac{dV}{dy} = E_y$$

$$-\frac{dV}{dz} = E_z$$

The units can be called newtons per coulomb or, equally well, volts per meter.

If we choose dl in the direction of the field, the value of dV/dl is a maximum and exactly equal to $|\mathbf{E}|$. We may write this as

$$-\left(\frac{dV}{dl}\right)_{\max} = |\mathbf{E}| \qquad (3.10)$$

This maximum rate of change of V at a given point is called the *gradient* of the potential at that point.

If we choose instead to move in a direction perpendicular to $\mathbf{E}$, $\cos\theta$ will be zero, and we find

$$\frac{dV}{dl} = 0 \qquad \text{or} \qquad V = \text{const}$$

The surfaces in space for which V = constant are called *equipotential* surfaces. From the above we see that they are everywhere perpendicular to the electric field direction.

Examples of the application of these ideas to particular cases are given in the next section.

3.7 Examples, Field versus Potential

a Field obtained from potential of a point charge Q This very simple example illustrates the method. Using Fig. 3.7, we write

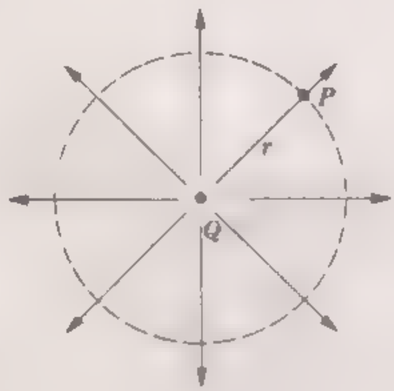


Fig. 3.7 The field at P due to a point charge Q can be obtained from $-dV/dr$.

directly from the result of Example 3.5*b* the potential at a point P a distance r from the charge Q:

$$V_P = \frac{1}{4\pi\epsilon_0}\frac{Q}{r}$$

Now we know the direction of the resultant field is along r, so we can write

$$|\mathbf{E}| = -\frac{dV}{dr} = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2} \qquad \text{volts/m}$$

in agreement with the earlier result.

Since when we move along r we are going in the direction of maximum rate of change of V, dV/dr is the gradient of V, or $-\mathbf{E}$.

Choosing instead to take the rate of change of V perpendicular to the radial direction, we find $dV/r\,d\theta = 0$, as we should expect, since the potential has a constant value as we move along a direction perpendicular to $\mathbf{E}$. Note that the elementary displacement along this direction perpendicular to $\mathbf{r}$ is given by $r\,d\theta$.

b Field of a dipole from the potential Using the relationship between potential gradient and electric fields, we are now able to find the field at any general point in the vicinity of a dipole. Using Fig.

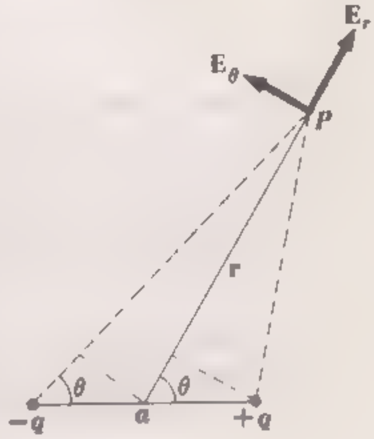


***Fig. 3.8** Calculation of potential and field around a dipole.*

3.8, we can write the expression for the potential at point P directly, using the expression for the potential due to point charges:

$$V_P = \frac{1}{4\pi\epsilon_0}\left\{\frac{q}{[r - (a/2)\cos\theta_1]} - \frac{q}{[r + (a/2)\cos\theta_2]}\right\}$$

$$= \frac{q}{4\pi\epsilon_0}\frac{a\cos\theta}{[r^2 - (a^2/4)\cos^2\theta]}$$

In the last expression we have made the approximation that $\theta = \theta_1 = \theta_2$. When P is reasonably far from the dipole, $r \gg a$, and we may neglect the term in $a^2/4$ compared with r^2. Under these conditions the approximation $\theta_1 = \theta_2$ is also valid. Using $p = qa$ for the dipole moment, we then have

$$V = \frac{1}{4\pi\epsilon_0}\frac{p}{r^2}\cos\theta \tag{3.11}$$

We do not have an explicit expression giving us the direction of $\mathbf{E}$, so we cannot obtain $\mathbf{E}$ from the gradient of V at once. We can get around this difficulty, however, by obtaining the components of $\mathbf{E}$ in two perpendicular directions. The magnitude and direction of the total field can then be obtained by simple vector addition of the two perpendicular components. Thus we can find

$$E_r = -\frac{\partial V}{\partial r} = \frac{1}{4\pi\epsilon_0}\frac{2p}{r^3}\cos\theta$$

and

$$E_\theta = -\frac{\partial V}{r\,\partial\theta} = \frac{1}{4\pi\epsilon_0}\frac{p}{r^3}\sin\theta$$

Comparison with our earlier results for $\theta = 0°$ and $90°$ shows agreement. Again, as in Example 3.7*a*, we have used $r\,d\theta$ for the elementary displacement perpendicular to $\mathbf{r}$.

3.8 Poisson's and Laplace's Equations

The inverse-square law puts very specific limitations on the way in which the electric field can vary from point to point in space. These limitations can be related to the continuous nature of lines of force or to Gauss' law. The exact formulation of these conditions is given by the equations of Laplace and Poisson, derived in Appendix A. The Laplace equation gives the way in which the potential must behave in a region containing no charge, and the Poisson equation does the same for regions in which there is a charge distribution.

3.9 The Evaluation of the Electronic Charge

Nothing in basic electric theory requires that there be a unique charge unit. Thus in our final formulation of the fundamental laws of electromagnetism in Chap. 12, we find no requirement that there be some minimum size of charge that cannot be subdivided. However, experiments with matter show that in fact the negative charge on the electron (or the equivalent positive charge on the proton) is indeed the basic unit of charge. No smaller subunits are known. In Chap. 6 we discuss experiments that evaluate the ratio of charge to mass, e/m, for various kinds of ions and for electrons. At the time of the discovery of the electron, however, there was no accurate method of evaluating m, so there was great need for an experiment that measures the electronic charge e directly. Efforts in this direction were brought to a successful conclusion in the period 1909–1913 by R. A. Millikan. In the experiment, very small oil drops were injected into the space between horizontal plates, across which an electric field could be varied by adjusting the voltage difference between the two plates. A microscope was arranged so that the vertical motion of the oil drops could be observed. An X-ray source was arranged so that X rays could be used to ionize the oil drops, giving them one or more units of charge. Two steps were required in the experiment. The first was to observe the motion of an oil drop charged by exposure to X rays and to adjust the electric field so that the vertical motion was arrested.

Under this circumstance, the electric force on a drop with n units of electronic charge, neE, just balances the difference between the gravitational force and the buoyant force of the air in which the drops are immersed. The net downward force is given by $\frac{4}{3}\pi r^3 g(\rho - \rho_a)$, where ρ is the density of the oil, g the acceleration of gravity, and ρ_a the air density. The radius r could not be measured accurately by observing its size in the microscope, since the drops were small (of the order of 10^{-4} cm), so a second step was taken to determine their size.

The method used was to turn off the electric field, by electrically connecting the two plates together, and to measure the terminal velocity of the downward-falling drop. Stokes' law for terminal velocity in a viscous medium gives

$$\frac{4}{3}\pi r^3 g(\rho - \rho_a) = 6\pi\eta r v \tag{3.12}$$

where η is the viscosity of the medium (air) and v is the terminal velocity. Thus after finding the electric field E for which a given drop was held in balance, so that

$$neE = \frac{4}{3}\pi r^3 g(\rho - \rho_a) \tag{3.13}$$

the radius of the drop was obtained by measuring its velocity in free fall according to Eq. (3.12).

By means of a large number of such observations, the total charge ne on many drops could be measured. Examination of the results showed that in all cases the charge on the drop was given by an integral number times a particular charge e, where

$$e = 1.602 \times 10^{-19} \text{ coulombs}$$

Thus, in the first place, the experiment showed that the electronic charge is unique; in the second place, it gave its value to great accuracy. The principal limitation on the accuracy of the experiment was the difficulty in the separate determination of the viscosity of air.

3.10 The Electron Volt

Since the potential difference V_{AB} gives the work per unit charge to move a charge between two points A and B, the energy gained by

a charged particle which is accelerated by the electric field between two points is given by

$$dU = qV_{AB} \qquad \text{coulomb-volts or joules} \tag{3.14}$$

In many situations a particle of charge equal to the charge e of an electron is involved, so a useful expression for energy is the *electron volt.* This is the energy gained by a particle of one electronic charge which is accelerated through one volt potential difference. Since the electronic charge is 1.602×10^{-19} coulomb, an electron volt is, from Eq. (3.14), 1.602×10^{-19} joule.

PROBLEMS

3.1 A dipole of charge $\pm q$ and separation l (dipole moment $p = ql$) is placed along the x axis as shown in Fig. P3.1.

Fig. P3.1

a Using the expression for the potential V of a point charge, calculate the work necessary to bring a charge $+Q$ from far away to a point S on the x axis, a distance a from the center of the dipole. What is the potential V_s of the point S (in the absence of the charge Q)?

b Write a simple approximate expression for V_s, good for $a \gg l$. Use the expression for V_s to find the magnitude and direction of the electric field at the point S.

c Find the orientation of the equipotential surface at the point S.

3.2 In Fig. P3.1, find an equipotential surface that is a plane. Find the value of the potential in this plane.

3.3 A uniform charge density of ρ coulombs/m^3 is in the shape of a sphere of radius R. Find expressions for the potential V and field E at distances r from the center, for points inside and outside the sphere.

3.4 Find the work necessary to move a charge q from a point A to a point B in the field of a point charge Q, as shown in Fig. P3.4.

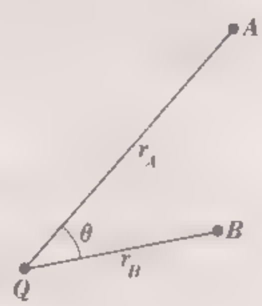


Fig. P3.4

3.5 The maximum electric field that can be supported in air (without producing ionization of the air and allowing charge to flow) is about 10^6 volts/cm. Using this criterion, find the maximum potential to which a conducting sphere of radius $R = 10$ cm can be charged in air.

3.6 A metal sphere has a radius R and is isolated from all other bodies. Express the potential of the surface of the sphere as a function of the charge placed on it. Integrate this expression to determine the work necessary to charge the sphere up to a potential V.

3.7 A spherical conductor of radius a has a charge Q_1 placed on it. This is surrounded by a thin spherical conducting shell of radius b, as shown in Fig. P3.7. The shell is connected to ground through a battery of potential difference V_1.

a Find the total charge on the outer surface of the shell and on the inner surface of the shell.

b Find the field and potential at a radius r from the center of the sphere, where $r < a$, $a < r < b$, $r > b$.

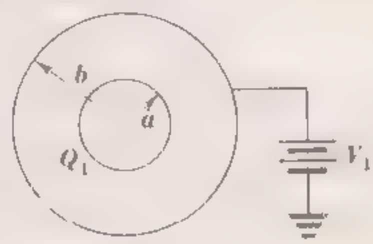


Fig. P3.7

3.8 What is the velocity of an electron that has been accelerated through a potential difference of 100 volts? What is its energy in joules? In ergs? In electron volts?

3.9 A long cylinder of radius a has a charge of Q coulombs/m. Find the potential difference between two points at distances r_1 and r_2 from the axis of the cylinder.

3.10 The graph of Fig. P3.10 shows the way in which the potential varies along the x axis. Plot a curve of the x component of the electric field E_x along the x axis. Explain why the two areas obtained for the E_x versus x plot should have equal magnitudes.

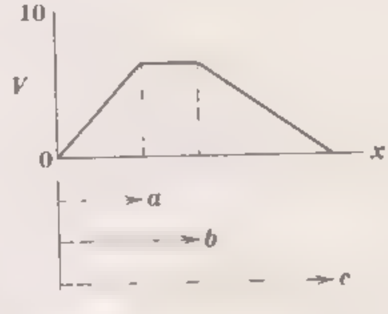


Fig. P3.10

3.11 Some of the electrons which are emitted at low velocities from a hot wire (cathode) go through a small hole in a plate that is at a potential

of 1,000 volts with respect to the cathode. What is the velocity of the electrons when they are passing through the hole? A second plate is parallel to the first and 20 cm beyond it and is at a potential of $-2{,}000$ volts with respect to the cathode. Describe the motion of electrons in the region between the two plates.

3.12 Electrons accelerated from rest through a potential difference of V_0 volts (having a kinetic energy of eV_0 electron volts) enter the middle of the vacuum region between two parallel plates of separation d and length b, as shown in Fig. P3.12. The potential difference between the two plates is V_1. Find the value of V_1 for which electrons just miss the edges of the plates. Assume $d \ll b$.

Fig. P3.12

3.13 Two identical water drops are charged to the same potential V_1. Find the new potential if the two drops coalesce into one drop.

3.14 Consider a point charge q a distance a from a conducting plane, as shown in Fig. P3.14. Since lines of force must end at the plane, charges are induced on the plane. We wish to find the force acting on q due to these induced charges. There is a very simple method for doing this if we invoke the uniqueness theorem mentioned in Appendix B, which states that any solution to an electrostatic problem that satisfies Laplace's equation (see Appendix A) and satisfies the boundary conditions is the only solution. Thus if we imagine an *image* charge of opposite sign placed symmetrically with respect to the conducting plane, as shown in the figure, the conducting plane corresponds to the zero potential equipotential surface of the dipole we have set up by using the image charge. Thus the field lines from the charge q are exactly those characteristic of a dipole of charge $\pm q$ and separation $2a$. Use this idea to calculate the force on q.

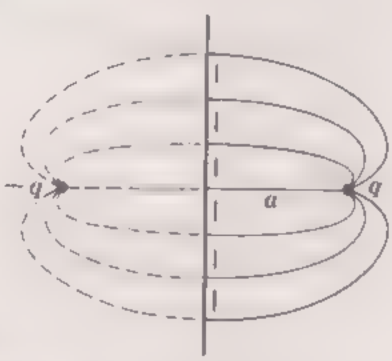


Fig. P3.14

3.15 Charges of $+\frac{4}{3} \times 10^{-8}$ coulomb and $-\frac{1}{3} \times 10^{-8}$ coulomb are placed along the x axis at the points -10 and 0 cm, respectively.

a Make a plot of the potential as a function of x at any point along the x axis and also as a function of position on a line perpendicular to the x axis and passing through the point $x = 10$ cm.

b At what points on the x axis is the potential 300 volts? Is the electric field intensity the same at these points?

c At what point would a third charge be in equilibrium? Would it be stable equilibrium?

3.16 The potential at points in a plane is given by

$$V = \frac{ax}{(x^2 + y^2)^{3/2}} + \frac{b}{(x^2 + y^2)^{1/2}}$$

where x and y are the rectangular coordinates of a point and a and b are constants. Find the components E_x and E_y of the electric intensity at any point.

3.17 The potential at points in a plane is given by

$$V = \frac{a \cos \theta}{r^2} + \frac{b}{r}$$

where r and θ are the polar coordinates of a point in the plane and a and b are constants. Find the components E_r and E_θ of the electric intensity at any point.

FOUR

Capacitance

4.1 Introduction

It is now clear that external work is needed to bring a group of charges together. This results from the forces of mutual repulsion between like charges. A case of considerable interest is that of the collection of excess charge on a conducting body or a group of conducting bodies. A quantity of importance here is called the *capacitance* of such a system. The capacitance of a conductor or a set of conductors is measured by the amount of charge that must be placed on it to raise its potential by 1 volt. Thus the capacitance C in *farads* is given by

$$C \text{ (farads)} = \frac{Q \text{ (coulombs)}}{V \text{ (volts)}} \tag{4.1}$$

The farad is the unit of capacitance in both the mks and the practical system of units (see Chap. 15). The larger the capacitance, the larger the amount of charge that must be added in order to raise the potential by 1 volt. The capacitance of a given geometric con-

figuration of conductors depends on the form of the electric field due to charges placed on the conductors, since the potential relates to the work to bring up charges from far away against the forces of the electric field.

We illustrate this connection between electric field and capacitance in the simple case of an isolated conducting sphere as in Fig. 4.1. Suppose we place a charge Q on a metal sphere. The field

***Fig. 4.1** Spherical conductor of radius r carrying a total charge Q.*

will be spherically symmetrical, and the potential at its surface will be given by

$$V = -\int_{\infty}^{r} \mathbf{E}\cdot d\mathbf{l} = \frac{1}{4\pi\epsilon_0}\frac{Q}{r}$$

where we have used for $\mathbf{E}$ the field of a point charge. Applying Eq. (4.1), we find $C = Q/V = 4\pi\epsilon_0 r$ farads. The capacitance of the isolated sphere increases linearly with the radius.

4.2 The Capacitor

Generally we have to deal with more complicated arrangements of conductors. Most often we are interested in a pair of conductors, which together are often called a *capacitor* (or *condenser*). Just as capacitance is of theoretical importance in the framework of electrostatic theory, the capacitor is one of the fundamental elements in electric circuits. It is particularly important in a-c and transient circuits. As an example of the capacitor we discuss the parallel-plate arrangement shown in Fig. 4.2. We shall connect the

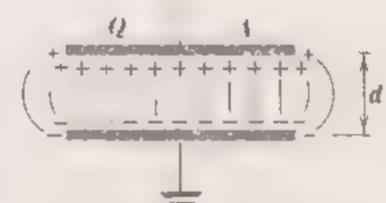


***Fig. 4.2** Parallel-plate capacitor, plate area A, separation d, charge density σ = Q/A.*

bottom plate to ground by a conducting wire. Since the earth is a relatively good conductor, as is the wire and plate, this fixes the plate at what we shall call zero potential.

When we place an excess charge, say positive, on the upper plate, if the upper plate is brought reasonably close to the lower, charges collect on the bottom surface of the upper plate and lines of force extend downward to the lower plate, where they terminate on negative charges. These are pulled up from the ground by the attraction of the positive charges. If the total charge on the upper plate is $+Q$, the total charge on the lower plate will be $-Q$. The proof that equal and opposite charges are pulled up from the ground to the bottom plate can be based on the ideas of lines of force. We know that lines of force from the positive charges leave the metal surface perpendicular to it. When the lines reach the lower plate, they must terminate on negative charges, since they cannot penetrate a conductor. In order that no lines penetrate the metal, the positive and negative charges on the two plates must be equal. It is the mutual attraction between unlike charges that brings them to the inner surfaces of the plates. As indicated in Fig. 4.2, there is essentially no field outside the region between the plates. This calculation neglects any small corrections due to *fringing* of the field at the edges of the plates.

The capacitance of this parallel-plate capacitor is calculated as follows. Let the separation between the plates be d, and let their area be A. Call the net charge on the upper plate Q (and therefore the lower plate holds a charge of $-Q$). The problem is to find the potential of the upper plate, with the lower plate held at $V = 0$, or, equivalently, the potential difference between the two plates. This we do by first finding the field in the region between the plates Then by integrating the field across the space between the plates, we obtain the potential difference. Using Gauss' law, as shown in Example 2.7c, we find that $E = \sigma/\epsilon_0$ or $E = Q/A\epsilon_0$. Since the field is uniform across the gap, the integration is simple.

$$V = -\int_0^d \mathbf{E}\cdot d\mathbf{x} = Q\frac{d}{\epsilon_0 A} \qquad \text{volts}$$

This gives for the capacitance $C = \dfrac{Q}{V} = \dfrac{A\epsilon_0}{d}$ farads.

Important aspects of this result include the fact that the capacitance of this capacitor is proportional to its area (probably

not surprising) and that the capacitance increases as the spacing between plates is made smaller.

The method used above shows that the calculation of capacitance involves integration of the electric field between two equipotential surfaces (usually conductors) to obtain the potential difference. This is then related to the charge responsible for the potential difference. From this point of view, the isolated sphere that was the first equipotential surface whose capacitance was studied may be considered as one electrode of a capacitor, the other electrode of which is at infinity. We have in effect assumed that the second electrode is so large and of such shape that the excess charge density at any point on it is trivial compared with the charge density on the sphere.

4.3 *Examples, Calculation of Capacitance*

a Capacitance between concentric spherical conductors Consider the outer sphere of radius b grounded as shown in Fig. 4.3. Place a

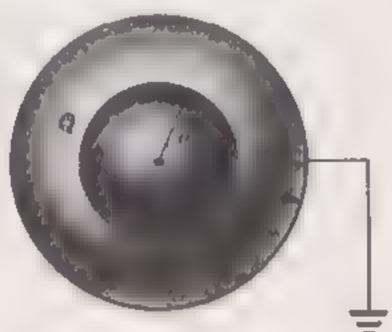


***Fig. 4.3** Two concentric metal spheres forming a capacitor.*

charge Q on the inner sphere (through a small hole in the outer sphere). Then an equal charge $-Q$ will be induced on the inner surface of the outer sphere (this follows from Gauss' law). Field exists only between the two spheres. The potential difference V_{ab} is given by integrating the field:

$$V_{ab} = -\int_b^a \mathbf{E}\cdot d\mathbf{r} = \frac{-Q}{4\pi\epsilon_0}\int_b^a \frac{dr}{r^2} = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{a} - \frac{1}{b}\right)$$

Then

$$C = \frac{Q}{V} = \frac{4\pi\epsilon_0}{1/a - 1/b} = 4\pi\epsilon_0 \frac{ab}{b-a} \qquad \text{farads}$$

Since we have already calculated the potential of a conducting sphere of radius r to be $V = \dfrac{1}{4\pi\epsilon_0}\dfrac{Q}{r}$, we can make this capacitance calculation alternatively by taking the difference between the potentials of the two spheres directly, instead of integrating the field over the distance between spheres. We would then write

$$C = \frac{Q}{V_{ab}} = \frac{Q}{(1/4\pi\epsilon_0)(Q/a - Q/b)} = 4\pi\epsilon_0 \frac{ab}{b - a} \qquad \text{farads}$$

It is left as a problem for the student to show that the capacitance between two coaxial cylinders is given by

$$C = \frac{2\pi\epsilon_0 L}{\ln (b/a)}$$

where a and b = radii of inner and outer cylinders
L = their length
ln = natural logarithm

b Capacitance of a conducting sphere surrounded by an isolated thick spherical conducting shell This rather artificial arrangement is useful to fix clearly the ideas of field and potential (Fig. 4.4). The thick

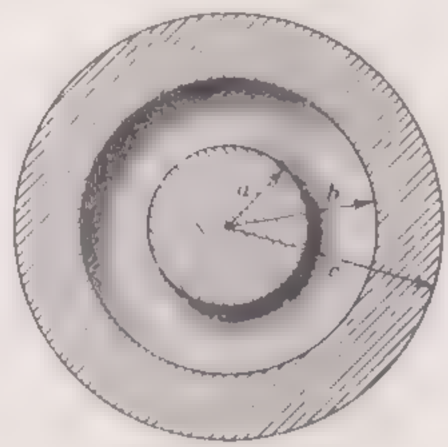


Fig. 4.4 Thick conducting shell surrounding a spherical conductor.

outer shell is isolated and considered to be initially uncharged. A charge $+Q$ is to be placed on the inner sphere. By Gauss' law we see at once that a negative charge $-Q$ is induced on the inner surface of the shell. This leaves a charge $+Q$ on the outer surface of the shell. Thus outside the metal shell the field is identical with the field due to a point charge $+Q$ at the center. We can therefore calculate the potential V_c at the surface by integrating the field from ∞ in the

usual manner. Within the metal shell (between c and b) the field must be zero, therefore V_b is the same as V_c (since $E_r = 0 = -\partial V/\partial r$). The field between b and a is also that of a point charge $+Q$, so the increase in potential at a over the value at b is obtained by integration from b to a. Then the potential at a, V_a, is given by

$$V_a = -\int_\infty^c \mathbf{E}\cdot d\mathbf{r} - \int_c^b \mathbf{E}\cdot d\mathbf{r} - \int_b^a \mathbf{E}\cdot d\mathbf{r}$$

$$= -\frac{Q}{4\pi\epsilon_0}\int_\infty^c \frac{dr}{r^2} - 0 - \frac{Q}{4\pi\epsilon_0}\int_b^a \frac{dr}{r^2}$$

$$= \frac{Q}{4\pi\epsilon_0 c} + \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{a} - \frac{1}{b}\right)$$

The capacitance can be obtained in the usual fashion.

4.4 Combinations of Capacitors

Capacitors are often combined in circuits. The two most common arrangements involve direct connections of the capacitors in *series* or in *parallel*. We give the simple arguments that allow calculation of the capacitance of such combinations. Figure 4.5 shows two

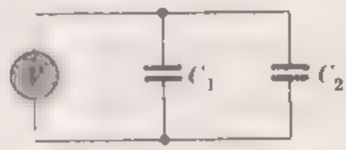


Fig. 4.5 *Two capacitors connected in parallel.*

capacitors connected in parallel. Since the electrodes are connected directly by conductors, the potential difference between the pairs of plates must be equal. The expression for this potential difference in terms of the capacitance and charge on each capacitor is

$$V_{ab} = \frac{Q_1}{C_1} \qquad V_{ab} = \frac{Q_2}{C_2}$$

The capacitance of the equivalent single capacitor is obtained by comparing the total charge $Q_1 + Q_2$ with the potential difference V_{ab}. Thus,

$$C = \frac{Q_1 + Q_2}{V_{ab}}$$

But since $C_1 = Q_1/V_{ab}$ and $C_2 = Q_2/V_{ab}$, $C = C_1 + C_2$. In general, for any number of capacitors in parallel, the resultant capacitance is given by

$$C = \sum_i C_i \tag{4.2}$$

Figure 4.6 shows three capacitors in series. We start off with no charge on any of the capacitors and then add a charge $+Q$ to the first capacitor. Equal and opposite charges will then be induced on consecutive plates. In contrast to the parallel case, which gave equal voltages across each capacitor, we find here that the charges are equal on each capacitor. Thus $Q = Q_1 = Q_2 = Q_3$. The total voltage across the set will be given by $V_1 + V_2 + V_3 = V$. But $V_1 = Q/C_1$, $V_2 = Q/C_2$, and $V_3 = Q/C_3$. Thus

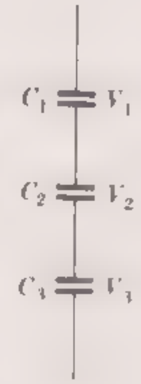


Fig. 4.6 Three capacitors in series.

$$V = Q\left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}\right),$$

and the equivalent capacitance is given by $\dfrac{1}{C} = \dfrac{1}{C_1} + \dfrac{1}{C_2} + \dfrac{1}{C_3}$ or, in general for series connections,

$$\frac{1}{C} = \sum_i \frac{1}{C_i} \tag{4.3}$$

That is, in series the reciprocal of the equivalent capacitance is the sum of the reciprocals of the capacitances.

4.5 Stored Energy in Capacitors

In Sec. 4.1 we pointed out the fairly obvious fact that whenever a group of charges are brought together, as on a capacitor, work has to be done against the repulsive forces between like charges. The work done is stored as potential energy. This follows from the absence of dissipative forces, or by virtue of the conservative na-

ture of the electrostatic field. In any case, it is of interest to calculate the energy stored in a charged capacitor. This is easy, provided we remember that as the capacitor gets charged to a higher and higher potential, the work to bring up each increment of charge increases, as the collected charge increases. We show in Fig. 4.7 a

Fig. 4.7 A charge dQ is added to a charge Q on a capacitor.

partially charged capacitor with a charge Q. An incremental charge dQ is being brought up against the forces of the field. In general, the work necessary to bring up an increment of charge dQ to a capacitor that is at a potential V will be given by $dU = V\,dQ$. Since V varies during the charging process, we must somehow change variables before integrating. A convenient method involves substituting $C\,dV$ for dQ. This relationship comes from $Q = CV$ by differentiating. The expression for the work to charge a capacitor then becomes

$$U = \int_0^V CV\,dV = \tfrac{1}{2}CV^2 \qquad \text{joules} \tag{4.4}$$

Using Eq. (4.1), we can express this also in the alternative forms,

$$U = \frac{1}{2}QV = \frac{1}{2}\frac{Q^2}{C} \qquad \text{joules}$$

Where is this energy stored? It is useful to consider the energy to be stored in the electric field. Thus we may consider that the work we do in charging the capacitor produces an electric field in space and that the existence of the field involves stored energy. Without insisting that this is the only possible point of view, we can determine the relationship between stored energy and magnitude of electric field. The problem is simplified if we examine the parallel-plate capacitor. This has the convenient property that the space in which the field exists is well defined (neglecting fringing effects); also, the field is uniform in the region in which it exists. Thus if we

divide the total work done in charging the capacitor by the volume in which there is a field, we may compare the work or stored energy per unit volume with the value of the electric field. The calculation follows.

Choose a parallel-plate capacitor of area A and plate separation d. Its capacitance, we have seen, is $C = \epsilon_0 A/d$. The work to place a charge Q on the plates is given by $U = \frac{1}{2}Q^2/C$. Substitution of the expression for C and use of the relationship $Q = \sigma A$ gives $U = \frac{1}{2}\sigma^2 Ad/\epsilon_0$. Since the volume containing the field is just Ad, we have $U/\mathrm{vol} = \frac{1}{2}\sigma^2/\epsilon_0$. Now we know that E due to a plane conductor with a surface charge density σ is σ/ϵ_0. Substitution gives

$$\frac{U}{\mathrm{vol}} = \frac{1}{2}\epsilon_0 E^2 \tag{4.5}$$

Thus the energy stored in a volume depends on E^2 in that volume. Even in the more usual situations where the field varies in magnitude and direction over the region in which it exists, the stored energy in each small volume element is proportional to E^2. Thus we can write the more general expression for the total stored energy in an electrostatic field as

$$U = \tfrac{1}{2}\epsilon_0 \int E^2\, dv \tag{4.6}$$

where the integration is over all space. However, the contribution to this integral from point charges is infinite, since E will go to infinity as we approach such a charge. Since real charges have a finite size, there is no real difficulty. The energy involved in the existence of charges is called the *self-energy* and is not included when we calculate the energy stored in an assembly of charges. This is allowable because the self-energy does not change when we change the relative positions of the charges.

Although in the case of a charged capacitor this assignment of the stored energy to the field is not necessary (we could equally well calculate the energy on the basis of the potential energy of the distribution of charges), it is a useful point of view, as is shown in Sec. 4.7. Moreover, as we shall see when we come to the discussion of the propagation of energy in electromagnetic waves, it becomes almost essential to take this point of view.

4.6 *Self-energy of Electric Charges*

The expression for the energy stored in an electric field derived above allows us to investigate the self-energy of electric charges in more detail. Suppose we have two isolated charges, say electrons, which are far apart. The self-energy of each electron, which results from the coulomb field of each charge, is given by Eq. (4.6). If the two charges are brought close enough together so that there is an appreciable overlap of the fields, the total stored energy will be given by the same expression, where E is the vector sum of the fields $\mathbf{E}_1$ and $\mathbf{E}_2$ from the individual charges. We may thus write for the total energy

$$U = \tfrac{1}{2}\epsilon_0 \int (\mathbf{E}_1 + \mathbf{E}_2)^2 \, dv$$

$$= \tfrac{1}{2}\epsilon_0 \int E_1{}^2 \, dv + \tfrac{1}{2}\epsilon_0 \int E_2{}^2 \, dv + \tfrac{1}{2}\epsilon_0 \int 2(\mathbf{E}_1 \cdot \mathbf{E}_2) \, dv$$

The first two terms are simply the self-energies of the two separate electrons, while the third term is the extra energy resulting from the overlapping of the two fields. It is possible to show that this third term is equal to $e^2/4\pi\epsilon_0 r$, where r is the separation of the two charges and e is the charge on each electron. This is just the usual expression for the potential energy of one charge in the field of the other. This result illustrates the fact that in calculating the work done to assemble a group of charges we neglect the self-energy terms.

4.7 *Force between Capacitor Plates*

The foregoing considerations of energy stored in an electric field allow us to make a simple calculation of the force between charged electrodes on the basis that the work done in changing the stored energy by moving an electrode must be accounted for by the product of force times distance. Thus where $\mathbf{F}$ is the attractive force between the charged plates of a capacitor, we must apply a force $-\mathbf{F}$ to move them apart (quasi-statically). The external work necessary to increase the plate separation an amount dx will be $-F\,dx$ (where dx is taken positive for increasing separation). The resulting increase in stored energy must be $dU = -F\,dx$. We apply this to the parallel-plate capacitor, of area A and plate separation x.

We choose the case of isolated electrodes so that the charge Q on them is constant. We then choose the form for stored energy that involves Q. $U = \dfrac{Q^2}{2C} = \dfrac{Q^2x}{2\epsilon_0 A}$. For a small displacement dx, we find $dU = (Q^2/2\epsilon_0 A)\,dx$. Since $dU = -F\,dx$, we find for this case that $F = -Q^2/2\epsilon_0 A$, where the negative sign indicates that $\mathbf{F}$ is in the negative x direction, giving attraction between plates.

This is one example of the very useful method for determining forces through energy-work considerations.

4.8 Solution of Potential Problems, General

In determining the capacitance of various capacitors, we have seen that the central problem is to determine the field and from this the potential due to certain charge configurations. A more general problem is to determine the potential or field at all points due to any configuration of equipotential surfaces. In general, this can be a very difficult mathematical task, though there are quite a number of special cases for which the mathematics is reasonably simple. The solution of these potential problems comes under the heading of *boundary-value* problems. Once the values of the potentials of the equipotential surfaces (the boundaries) are fixed (including perhaps that the potential at infinity goes to zero), the problem can in principle be solved. Solution is greatly simplified by a *uniqueness* theorem, which states that the problem has only one solution, so *any* solution that fits the given boundary conditions is the *only* solution. To prevent being too much diverted by mathematical detail, we postpone for later study any general examination of boundary-value potential problems. Appendix B deals with one simple case as an example of such problems.

4.9 Some Electrostatic Apparatus

The electroscope, potential measurement We have already seen in Chap. 1 that the separation of the foils of an electroscope results from the mutual repulsion of like charges placed on the foils. We can now see that the amount of charge will depend on the potential placed across the electroscope, since the potential of the foils will depend on the amount of charge on them. Thus the electroscope gives a direct measure of potential. The electroscope and similar

electrostatic devices are particularly useful for measuring potential in certain applications since no current flow is required.

It is also possible to use the electroscope for determining small currents. Suppose the charge on a capacitor is leaking off very slowly (i.e., a very small current is flowing). An electroscope attached to the capacitor will measure the rate at which the potential decreases. If the total capacitance of the system is C and the original charge is Q, we have from Eq. (4.1) that the potential is given by $V = Q/C$. The rate of change of charge on the system can be related to the rate of change of potential by differentiating this equation, taking C as a constant. Thus we find

$$\frac{dV}{dt} = \frac{1}{C}\frac{dQ}{dt} = \frac{1}{C}i \tag{4.7}$$

Here we have written i, the *current*, for dQ/dt. (We shall discuss current in Chap. 7.) Thus if we know or can measure the capacitance of a system, measurement of dV/dt on an electroscope connected across the system will allow measurement of current.

Kelvin's absolute electrometer The electrostatic force between charged parallel plates is used in an absolute electrometer as devised by Kelvin for measuring potential difference in absolute units. A pair of plates are connected as shown in Fig. 4.8, with the upper

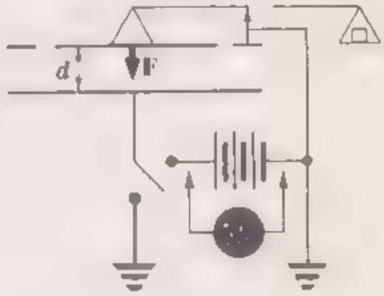


***Fig. 4.8** Kelvin's absolute electrometer. Note use of guard ring to eliminate effects of fringing of field at edge of capacitor.*

plate connected to a balance. An annular *guard ring* surrounds the circular upper plate. Its function is to keep the field uniform in the central region between the plates to avoid unwanted edge effects. There is fringing or bulging of the field lines at the outer radius of the guard ring, but this does not affect the field around the upper plate. After the balance is adjusted until the upper plate is just in the plane of the guard ring with no charge on the system (lower plate grounded), the lower plate is connected to the potential to be measured (shown as batteries in the sketch). Additional weights

mg are then added to the balance until the upper plate is returned to its original position. Then, from Sec. 4.7, we have $F = mg = Q^2/2\epsilon_0 A$. Since $Q/A = \sigma$, $E = \sigma/\epsilon_0$, and $V = Ed$, we find

$$mg = \frac{1}{2}\frac{\epsilon_0 A}{d^2}V^2 \qquad \text{or} \qquad V = d\sqrt{\frac{2\,mg}{\epsilon_0 A}} \qquad \text{volts}$$

PROBLEMS

4.1 Find the capacitance of a pair of coaxial metal cylinders of radii a and b and length l, as shown in Fig. P4.1.

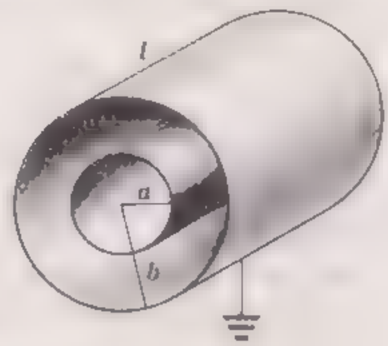


Fig. P4.1

4.2 Figure P4.2 shows a metal sphere of radius a, surrounded by a spherical thick metal shell of inner and outer radii b and c. This shell is isolated electrically, with no net charge, and is surrounded by a grounded spherical shell of radius d. Find the potential of the inner sphere when a charge Q is placed on it. What is the capacitance of the sphere?

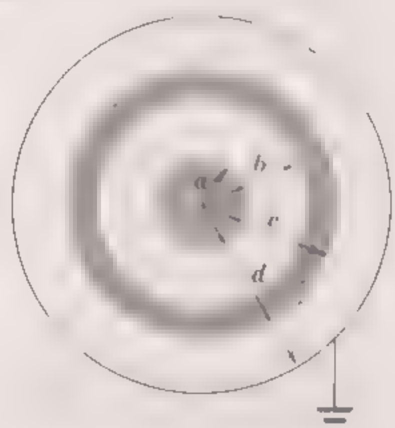


Fig. P4.2

4.3 Find the capacitance of the combination of capacitors shown in Fig P4.3.

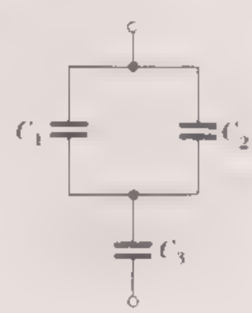


Fig. P4.3

4.4 Find the capacitance of the combination of capacitors shown in Fig. P4.4.

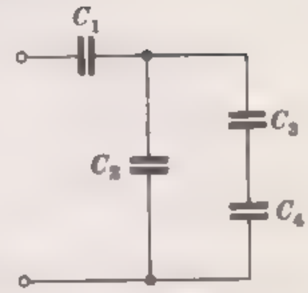


Fig. P4.4

4.5 A parallel-plate capacitor with plate separation d has a capacitance C_1. Find its new capacitance when an isolated metal slab of thickness a is placed between the plates.

4.6 A parallel-plate capacitor of plate separation d is charged to a potential difference V_1 and isolated. The plate separation is increased to $2d$. What is the new potential V_2 between the plates? By how much is the energy stored in the capacitor increased? Where did this energy come from?

4.7 Find the force of attraction between two parallel metal plates of area A separated by a distance d and charged to a potential difference V. Do this by calculating the force on a charge dq on one plate due to the field of the charges on the opposite plate. This can then be integrated to give the total force on all charges on the first plate.

4.8 In the arrangement shown in Fig. P4.8, find the necessary relationship between the capacitances of the four capacitors in order that when a voltage is applied across terminals a and b, no voltage difference is set up between terminals c and d. Would this arrangement work for the case of the voltage applied to terminals c and d to give no voltage across a and b?

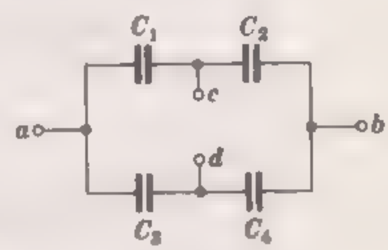


Fig. P4.8

4.9 Calculate the capacitance of the multiplate capacitor shown in Fig. P4.9. The area of overlap of the plates is A and the separation between consecutive plates is d. Tuning capacitors for radios, etc., are made in this fashion, by a mechanical arrangement that allows the area of overlap to be varied.

Fig. P4.9

4.10 What potential would be necessary between the parallel plates of a capacitor separated by a gap of 1 cm in order that the gravitational force on an electron would be balanced by an upward electric field? What potential would be required to balance the gravitational force on a proton?

4.11 A potential difference of 200 volts is applied across a 2-μf and a 6-μf capacitor connected in series. What is the potential difference across each capacitor, and the charge on each?

4.12 Find the capacitance of the earth (radius 4,000 miles).

4.13 An isolated conducting sphere of radius R has a charge Q. What is the total stored energy? What is the radius r within which half the stored energy is contained?

4.14 A sphere whose radius is 0.2 m is charged to a potential of 30,000 volts.

a What is its stored energy?

b If it is connected by a very long wire to an identical uncharged sphere located at a very great distance, what is the final energy of this system?

4.15 A spherical capacitor has radii of inner and outer spheres a and b, respectively. The inner sphere bears a charge q. What total charge must be placed on the outer ($r = b$) sphere in order to confine the electric field to the space between the spheres $a < r < b$?

4.16 A crude idea of the size of an electron can be obtained from the following model. Assume the electron is a sphere of radius a, with its charge distributed uniformly over its surface. Calculate the total electrostatic energy involved in this charge distribution and equate this to mc^2, where m is the mass of the electron and c is the velocity of light (3.0×10^8 m/sec). Use this to calculate the radius of the electron.

4.17 Two capacitors, one charged and the other uncharged, are connected in parallel. Prove that when equilibrium is reached, each capacitor carries a fraction of the initial charge equal to the ratio of its capacitance to the sum of the two capacitances. Show that the final energy is less than the initial energy, and derive a formula for the difference in terms of the initial charge and the capacitances of the two capacitors.

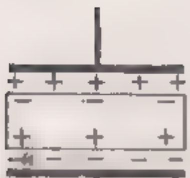
4.18 A parallel-plate capacitor has plates of area 500 cm^2, separated by a distance of 1.0 cm. A potential difference of 2,000 volts is applied between the plates, after which they are isolated.

a What is the energy stored in the capacitor?

b An uncharged sheet of metal 2.0 mm thick is placed between the plates and parallel to them. How much work is done by electric forces during the insertion of the metal sheet?

c What is the potential difference between the capacitor plates after the sheet has been inserted?

FIVE



Dielectrics

5.1 Introduction

We have been considering the problems of electrostatics in the virtual absence of matter, except for the use of conductors to establish equipotential surfaces. We now begin our investigation of the effects of the presence of matter. In what would now be classed as an experiment in solid-state physics, Faraday in 1837 repeated independently the earlier experiments of Cavendish (in about 1770) showing that when the space between the plates of a capacitor is completely filled with insulating matter such as glass or mica, the capacitance is multiplied by a factor K greater than 1. This factor, which is called the *dielectric constant* or *specific inductive capacitance*, is independent of the shape and size of the capacitor, but its value varies widely for different materials. Free space has the value 1 (by definition), various kinds of glass have values around 6, water has the value 81, and the value for air is 1.0006, so close to 1 that we ordinarily neglect its effect. All materials that are not conductors show this effect and are called *dielectrics*. Our purpose in

this chapter is to find out the causes of this effect and to determine the additional parameters that describe it.

5.2 Polarization of Matter

The ultimate source of dielectric behavior is the electrical nature of matter. Although normally electrically neutral as a whole, in detail matter is made up of positive and negative charges in equal quantity. In dielectric materials these charges are not free to move far under the influence of an external electric field, as are the conduction electrons in a conductor. However, the forces of an external field do cause small relative displacements (on an atomic scale) of charges of each sign. The extent of such motion depends on the tightness with which the charges are held fixed. The displacement of charge resulting from an applied external field is called *polarization* of the material.

We may take as a crude model of a dielectric two interpenetrating arrays of charge, as shown in Fig. 5.1, where the repre-

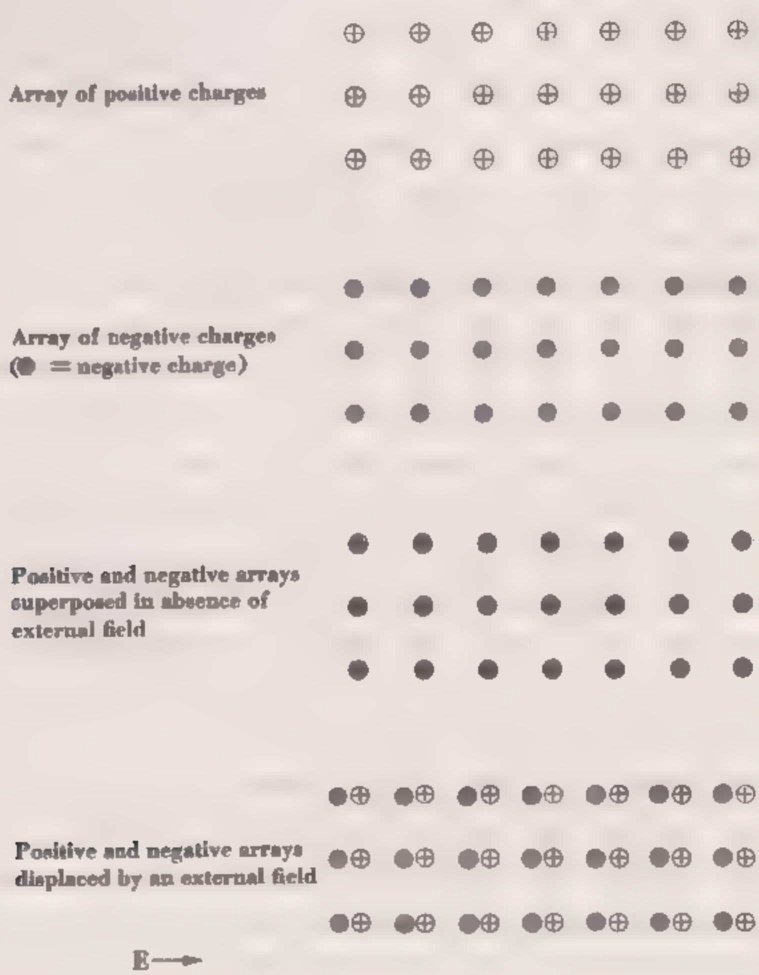


Fig. 5.1 Simple model of polarization.

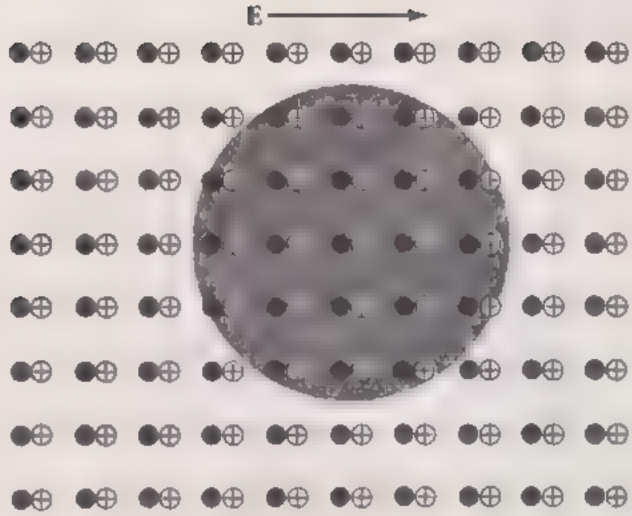
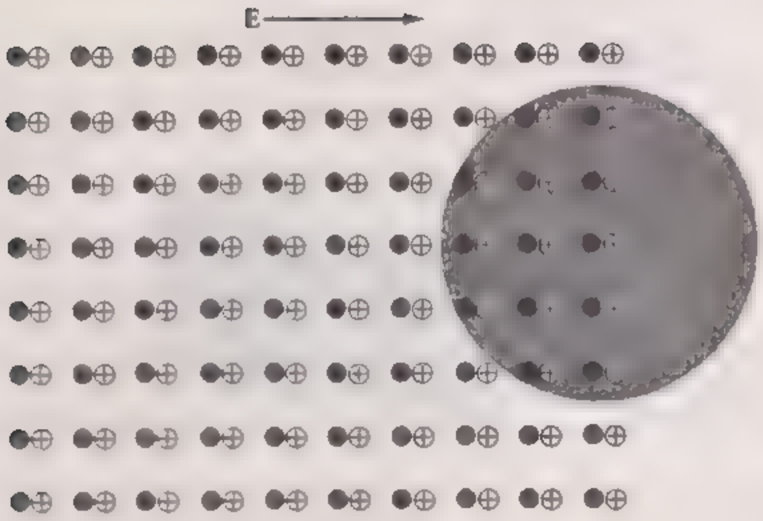


***Fig. 5.2** A polarized dielectric: (a) Volume from inside polarized matter has net charge = 0; (b) volume containing a surface normal to polarization direction has net induced surface charge.*

sentation is in two dimensions for simplicity. We can think of the positive array as being due to the nuclei of the atoms making up the solid, and the negative array as due to the average positions of the electrons associated with the atoms. In the absence of an external field we would expect the two arrays to be superposed, but the relative motion of the charges under an applied field produces a separation of charge as shown. This model is too crude to give an exact picture of the conditions inside the matter on a subatomic or atomic scale. It does, however, give a picture that is useful macroscopically. That is, the effect of the polarization on the field outside the matter is independent of the exact details of the model. Similarly, the *average* field inside the matter resulting from both the external field and the effects of polarization of the matter does not depend on the details of the model. The reason for this is easily seen. Take a small volume inside the body of the polarized material. Let this volume be enough to include a very large number of atoms. It could still be very small. (There usually are about 10^{22} atoms per cubic centimeter in a solid, so a volume containing 10^8 atoms would be 10^{-14} cm^3.) As suggested in Fig. 5.2*a*, the average charge in such a volume would be zero, despite the relative displacement of the charge arrays. For example, each positive charge has an equal negative charge to its left and to its right. On the other hand, for a volume including a boundary perpendicular to the direction of polarization (Fig. 5.2*b*), there is a net positive charge at the surface that is not compensated as are charges inside the material. From the macroscopic view, then, the important result of the charge dis-



(b)

placement due to polarization is the appearance of net charge at the surface of the matter. This is called *polarization charge.* In the absence of polarization it is compensated by superposed equal and opposite charge.

From the atomic point of view the relative displacement of charge centers produces another effect. The charge arrays can now be considered as a collection of electric dipoles. We can, in fact, speak of *the dipole moment per unit volume*, **P**. This is called the *polarization.* Notice that the word polarization has two meanirgs: a qualitative one, referring to any relative displacement of positive and negative charge, and a quantitative one, giving the resulting dipole moment of a unit volume of polarized matter. **P** may be regarded as a vector quantity having the resultant direction of the induced dipoles. Since the dipole moment per unit volume refers to the arrangement on an atomic scale, it is of fundamental interest in terms of understanding the nature of the solid. As shown above, however, it is not measurable directly by a macroscopic method. On the other hand, there is a simple relationship between **P** and the polarization charge on the surface of a dielectric slab. The effect of the surface charge is measurable directly and accounts for all the external effects of dielectrics. We show this relationship below.[1]

[1] When the dielectric material is not homogeneous, so that the dipoles induced are not uniform, there is an additional effect from a volume distribution of polarization charges, ρ_p. This charge density represents incomplete cancellation of the ends of the individual dipoles. We shall not discuss this complication further.

5.3 Polarization Charge versus Dipole Moment per Unit Volume

Figure 5.3 is an idealized picture of a block of dielectric placed in a uniform external electric field. The net surface charges resulting from the relative displacement of positive and negative charges

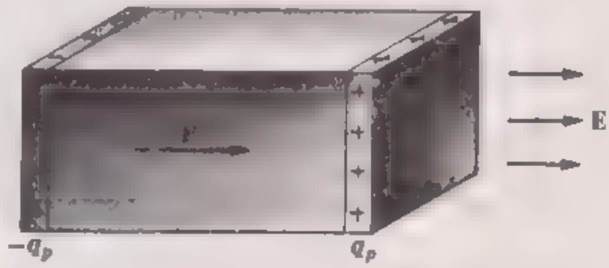


***Fig. 5.3** A polarized dielectric block showing net induced charge on surfaces.*

are shown. We shall call the density of polarization charge σ_p coulombs/m^2. The relationship between P and σ_p is obtained by computing the dipole moment of this block of dielectric in two ways. Thus the total dipole moment of the block is P times the volume of the block or PAL. The second way is to look on the block as one large dipole with surface charges $\pm Q = \sigma_p A$ separated by a distance L. The dipole moment of the block is then $QL = \sigma_p AL$. Comparison of the two results shows that

$$P = |\sigma_p| \tag{5.1}$$

The sign of the polarization surface charge depends on the direction of polarization and on which end of the block we consider. The magnitude of **P**, the dipole moment per unit volume, depends on the charge density and on the relative displacement of the + and − charge arrays. The direction of the vector **P** is along the displacement direction.

We can generalize Eq. (5.1) by taking a situation in which the dielectric surface is not perpendicular to **P** (Fig. 5.4), where the

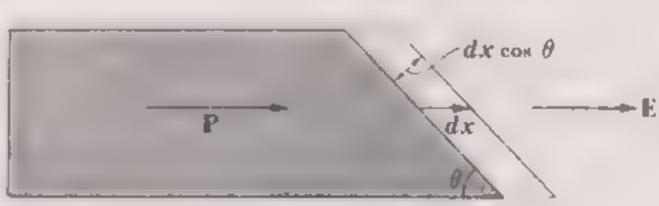


Fig. 5.4** Surface polarization charge related to polarization* **P. Actual relative displacement of positive charge is dx; effective displacement is dx cos θ.*

normal to the surface makes an angle θ with the direction of **P**. In this case the actual displacement of the array of positive charges within the dielectric is dx. However, since the displacement is canted with respect to the surface, the effective displacement of the charges

is only $dx \cos \theta$. Thus for surfaces parallel to **P**, no surface polarization charge is induced and, in general,

$$P_n = -\sigma_p \tag{5.2}$$

where P_n is the component of **P** normal to the surface.

We have inserted a negative sign in Eq. (5.2). The explanation of this is as follows: The polarization **P** is a vector field quantity having mathematical properties very similar to **E**. For uniformly polarized matter it can be represented by lines that originate and end on surface polarization charges, and Gauss' law applies to these lines. But the direction of lines of **P** is from negative to positive charge, so lines of **P** emerge from a volume containing negative polarization charge and end in a volume containing positive polarization charge. It follows that

$$\int_{CS} \mathbf{P} \cdot d\mathbf{S} = -q_p$$

This equation solved for P_n gives rise to Eq. (5.2) with the negative sign as shown. Of course **P** can exist only in regions occupied by matter. Thus the macroscopic effect of the production of atomic dipoles manifests itself outside by the field of the polarization surface charge. That is, when a block of matter is polarized, we can explain the macroscopic effects on the field in its vicinity by replacing the entire block with the resulting surface polarization charges. In doing this, we are ignoring the particulate nature of matter and considering it to be made up of two superposed continua of uniform positive and negative charge, as rather suggested by Fig. 5.3. We need consider the particulate nature of matter only when we wish to know the way in which the field varies within the matter on an atomic scale, or to relate the value of K to the atomic structure.

Let us now make the additional simplifying assumption that the polarization is proportional to the applied field. With this assumption we can write

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E} \tag{5.3}$$

where χ is called the *electric susceptibility* and **E** is the *macroscopic* field (i.e., the average field due to the external applied field *as modified by* polarization surface charges). This assumption that the polarization depends linearly on the field corresponds to the be-

havior of most materials. The susceptibility of a material relates its polarization to the macroscopic field causing the polarization. The constant ϵ_0 is included only for the purpose of simplifying the form of later relationships.

Now that we have seen that the macroscopic electric field due to polarized matter can be discussed in terms of the polarization surface charge, it is possible to write a general expression relating the electric field to the free charge and the polarization surface charge. This is simply Gauss' law:

$$\int_{CS} \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0}(q_f + q_p) = \frac{1}{\epsilon_0} q_{\text{total}} \tag{5.4}$$

which states that the flux of electric field lines out of a given volume depends only on the total charge, both free and polarization, inside that volume. This is a very important general relationship, but we must realize that it is not necessarily easy to go from this equation to a solution for the electric field configuration. We next apply this equation, which relates to the sources of electric field, to a simple geometric arrangement and show how it can be used to obtain a detailed knowledge of the field.

We examine the effect of placing a dielectric between the plates of a simple parallel-plate capacitor. Figure 5.5*a* shows the capacitor in a vacuum. We have called the mobile or free charge density

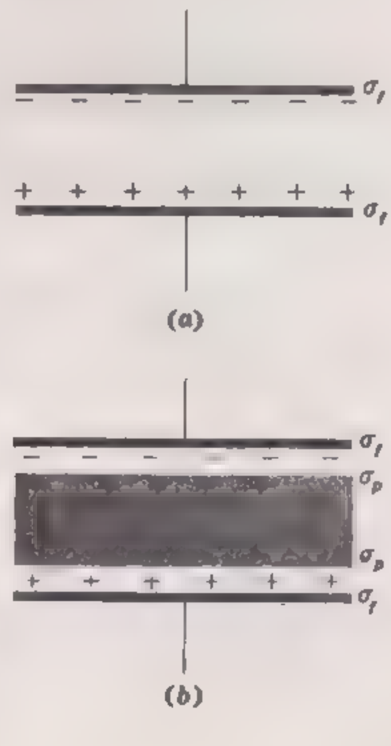


Fig. 5.5 Effect of dielectric in capacitor: (a) Free charge density σ_f; (b) polarization charge density σ_p adds algebraically to effect of σ_f.

σ_f. The uniform field between the plates, as we have seen, has a value $E = \sigma_f/\epsilon_0$.

In Fig. 5.5b the dielectric has been inserted so as to fill the entire space between the plates. As a result of the field originating from the free charges, the dielectric is polarized and there appears a polarization charge of density σ_p on the two surfaces. To calculate the average field within the dielectric, we must take account of both σ_f and σ_p. We apply Eq. (5.3) to the plane distribution of charge on the plates of the capacitor and on the surfaces of the dielectric sample. Since the electric field is uniform out from a plane charge distribution, we can take **E** out of the integral and solve for it. This gives

$$E = \frac{1}{\epsilon_0}(\sigma_f + \sigma_p) \tag{5.5}$$

where σ_f and σ_p are charge densities, obtained by dividing the total surface charge by the area of the Gaussian surface, which we here take to be the area of the plates. This is just like the procedure for the vacuum capacitor except for the inclusion of the polarization surface charge. Since σ_p is always of opposite sign to σ_f, the effect of the polarization surface charge is always to weaken the field as compared with the field in a vacuum. We discuss this in detail in Sec. 5.9, where we study the depolarization field.

The field E, as used in Eq. (5.5) to describe the situation in a dielectric, is actually only an *average* field. As we move from point to point on an atomic scale, the field varies widely in a periodic fashion. E here is simply the average value, which, when integrated, gives the true potential difference between the two plates of the capacitor.

We now calculate the capacitance of this capacitor filled with dielectric material, using, as in the vacuum case, $C = Q/V$. In this calculation we use for Q only the free charge, since we are concerned only with that charge which we can put on the plates. The polarization charge is, in effect, built into the dielectric material and so is not under our direct control. The capacitance equation can then be written as

$$C = \frac{\sigma_f A}{Ed} \tag{5.6}$$

where $\sigma_f A$ is the free charge on each plate. Since E is no longer related to σ_f by the vacuum equation $\sigma_f = \epsilon_0 \mathbf{E}$ because of the perturbing effects of σ_p, we find σ_f using Eq. (5.5), which takes account of the polarization charge. This gives $\sigma_f = \epsilon_0 E - \sigma_p$. If the macroscopic properties of the dielectric are characterized by its susceptibility [Eq. (5.3)], $-\sigma_p$ can be replaced by $\mathbf{P} = \epsilon_0 \chi \mathbf{E}$. This gives

$$\sigma_f = \epsilon_0(1 + \chi)E \tag{5.7}$$

This equation substituted into the capacitance equation gives

$$C = \frac{\epsilon_0(1 + \chi)A}{d} \tag{5.8}$$

Comparing this with the earlier result for the capacitance of a parallel-plate capacitor in vacuum, $C = \epsilon_0 A/d$, we see that $(1 + \chi)$ is the multiplicative constant K found by Faraday. Thus the dielectric constant is related to the susceptibility by

$$K = 1 + \chi \tag{5.9}$$

Another way of writing Eq. (5.8) is $C = \epsilon A/d$, where

$$\epsilon = \epsilon_0(1 + \chi) \tag{5.10}$$

ϵ is called the permittivity of the material. The dielectric constant may now be written as

$$K = \frac{\epsilon}{\epsilon_0} \tag{5.11}$$

the ratio of the permittivity of the material to that of free space. Since the dielectric constant K is defined as the ratio of the capacitance of a capacitor with and without a given material filling the space between electrodes, it has the same value in all systems of units.

There are two important things to note in this discussion of alternative ways of describing the dielectric properties of matter. One is that we have not yet related these properties to the polarizability of the individual atoms. The second is that our discussion involves filling the entire space in a capacitor with dielectric material. We still have to ask for the effect of placing a piece of dielectric material in, say, an extended uniform electric field. However, whatever the situation, the quantities P, K, χ, and ϵ are still valid parameters that describe the behavior of the material.

5.4 *Electric Displacement* D

A very useful new quantity that gives added insight into problems involving fields in the presence of dielectric media is the *electric displacement* **D**, a vector quantity that has some similarity to **E**. We begin by looking for a quantity that depends only on the free charges, in contrast to **E**, which depends on both free and polarization charges. The quantity we need can be defined by the vector equation,

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{5.12}$$

We shall show that this definition makes **D** a quantity which obeys Gauss' law but for which the sources are free charges only. Even with matter present, polarization charges are not involved in **D**. We justify this defintion of **D**, and in fact the entire use of **D**, on the basis that it is a convenient tool for solving a multitude of problems. In this respect and a number of others, **D** is similar to **E**.

We begin by substituting for **E** from Eq. (5.12) in the general equation (5.4). This gives us

$$\epsilon_0 \int_{CS} \mathbf{E} \cdot d\mathbf{S} = \int_{CS} \mathbf{D} \cdot d\mathbf{S} - \int_{CS} \mathbf{P} \cdot d\mathbf{S} = q_{\text{total}} \tag{5.13}$$

We can use this form of the general equation to determine the properties of **D**. We need only to evaluate the integral containing **P**, which can be done on the basis of our earlier investigation of the polarization surface charge at a dielectric boundary in which we found $P_n = -\sigma_p$. We see at once that if **P** is uniform, $\int_{CS} \mathbf{P} \cdot d\mathbf{S} = 0$ anywhere inside a dielectric body. To show this, we take a Gaussian volume such as A in Fig. 5.6, with plane surfaces S perpendicular

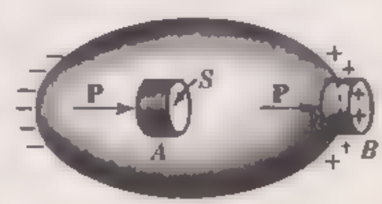


***Fig. 5.6** Gaussian volumes inside and at surface of a polarized dielectric body.*

to **P**. Clearly, PS at the left-hand surface is the negative of PS at the right-hand surface, and the integral over the curved surface is zero, so the integral over the entire closed surface must be zero. When, however, the Gaussian surface encloses a surface of the dielectric that has a component of **P** perpendicular to it, as at B in

the figure, the integral on the left-hand surface has a magnitude P_nS and on the right, outside the dielectric, has a value zero. Thus we have

$$\int_{CS} \mathbf{P}\cdot d\mathbf{S} = \int_{CS} P_n\, dS = -\sigma_p\, dS = -q_p \tag{5.14}$$

where q_p is the total polarization charge inside the Gaussian surface. The negative sign results from the fact that **P** points *inward* to the Gaussian volume when the polarization charge is positive, and *outward* when the charge is negative. This is just opposite to the case of the field **E**.

Now that we have evaluated $\int_{CS} \mathbf{P}\cdot d\mathbf{S}$, we can solve in Eq. (5.13) for $\int_{CS} \mathbf{D}\cdot d\mathbf{S}$. We find

$$\int_{CS} \mathbf{D}\cdot d\mathbf{S} = q_{\text{total}} - q_p = q_f \tag{5.15}$$

Thus the vector quantity **D**, which we have invented by writing Eq. (5.12), obeys Gauss' law; the sources are free charges only, even when dielectric material is present. In regions outside dielectric materials, where only free charges are present, $\epsilon_0 \int_{CS} \mathbf{E}\cdot d\mathbf{S} = q_f$ and $\int_{CS} \mathbf{D}\cdot d\mathbf{S} = q_f$, so

$$\mathbf{D} = \epsilon_0 \mathbf{E} \tag{5.16}$$

Thus D and E outside of matter have the same direction and differ in magnitude only by the constant ϵ_0.

This new concept of displacement is now applied to the simple problem of a dielectric slab between the plates of a parallel-plate capacitor (Fig. 5.7). We have seen already that $E = (\sigma_f + \sigma_p)/\epsilon_0$

Fig. 5.7. Dielectric slab between capacitor plates.

in this case. This may be rewritten, according to Eqs. (5.3) and (5.10), as $E = (1/\epsilon)\sigma_f$. We now apply Gauss' law for **D** as given in Eq. (5.15) and find

$$D = \sigma_f \tag{5.17}$$

Comparison of these equations for **D** and **E** shows that

$$\mathbf{D} = \epsilon\mathbf{E} \tag{5.18}$$

in matter. Thus if we know the free charge density on the capacitor plates and ϵ for the material, we may find **E** by first obtaining $D = \sigma_f$ and then using $\mathbf{D} = \epsilon\mathbf{E}$.

Before going to some other illustrative examples to help fix these new ideas, we give an additional discussion of **E** and **D** that is sometimes helpful. Imagine again a parallel-plate capacitor filled with a dielectric. Let there be two cavities cut inside the material as shown in Fig. 5.8. The left-hand cavity is coin-shaped, oriented

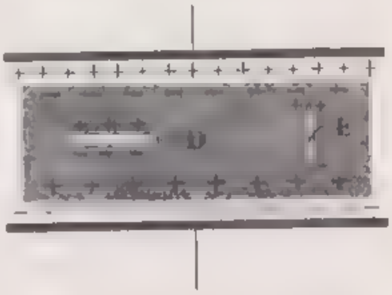


***Fig. 5.8** Cavities in a dielectric slab between capacitor plates. The value of **D** in the left-hand coin-shaped cavity is the same as though no cavity were cut. Similarly, the value of **E** in the right-hand needle-shaped cavity is the average **E** in the dielectric material.*

with its faces perpendicular to the field, and the right-hand cavity is a long needle-shaped cylinder, oriented along the field direction. We shall determine the field in the middle of these two contrasting cavities. In the coin-shaped cavity the field results only from the charges q_f. This follows from the cancellation of the fields of the polarization charges on the outer surface of the dielectric and those of opposite sign on the surface of the cavity. Since these surface charges are both plane distributions having the same charge density, their equal and opposite uniform fields cancel, leaving only the field of the free charge on the capacitor plates. Thus the field inside the coin-shaped cavity is $E = \sigma_f/\epsilon_0$, or

$$\mathbf{D} = \epsilon_0\mathbf{E} = \sigma_f \tag{5.19}$$

This value of **D** in the free space of the cavity is just the average value of **D** inside the dielectric if there were no cavity.

In the needle cavity, if it is long and thin enough, we may neglect the effect in the middle of the cavity due to the small cluster of polarization charges at the ends, so the field inside this cavity is given by

$$E = \frac{1}{\epsilon_0}(\sigma_f + \sigma_p)$$

This corresponds just to the average **E** inside the dielectric in the absence of a cavity, as found in Eq. (5.5).

We have arrived here, through a particular example, at an important pair of boundary conditions, for **E** and for **D**, which we discuss later in more detail. If we generalize the results above, we have: (1) At the boundary of a dielectric, the component of **D** normal to the surface is continuous. (Thus **D** in the coin-shaped cavity is normal to the surface and is indeed of the same value inside and outside the cavity.) (2) At the boundary of a dielectric, the component of **E** parallel to the surface is continuous. (In the needle-shaped cavity **E** is parallel to the surface and equal to the value in the body of the dielectric.)

5.5 Examples

a Electric field in a parallel-plate capacitor filled with dielectric We discuss this simple case to show the contrast between the use of **D** and **E**. It was shown in Sec. 5.4 that when Gauss' law is applied to the parallel-plate capacitor, the electric field is given by $E = (\sigma_f + \sigma_p)/\epsilon_0$. Substitution for σ_p in terms of ϵ or K gave

$$E = \frac{1}{\epsilon}\sigma_f = \frac{1}{K\epsilon_0}\sigma_f$$

The alternative approach is to apply Gauss' law to the displacement vector **D** [Eq. (5.15)]. This gives

$$DA = q_f \qquad \text{or} \qquad D = \frac{q_f}{A} = \sigma_f$$

where $A = \int d\mathbf{S}$, the flat surface of the Gaussian pillbox through

which the lines of D emerge. Since $\mathbf{D} = \epsilon\mathbf{E}$ inside the dielectric, the last equation can be written as

$$E = \frac{1}{\epsilon}\sigma_f$$

in agreement with the result obtained through **E** above. Thus **E** may be obtained by expressing **E** in terms of both σ_f and σ_p; or, by dealing with **D**, the same result may be obtained without explicit use of σ_p. The capacitance of the capacitor can be obtained as before, by substitution of the value of **E** in Eq. (5.6).

b Capacitance of a capacitor filled with two different dielectric slabs Let the dielectric constants and thicknesses of the two slabs be K_1 and a, and K_2 and b (Fig. 5.9). The lines of **D** are continuous

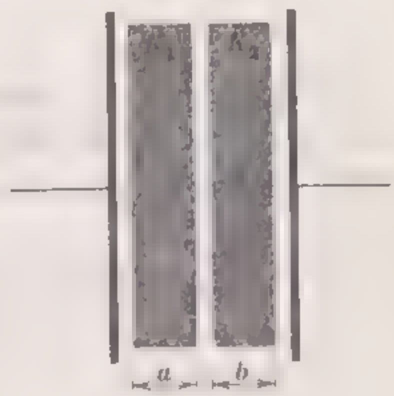


***Fig. 5.9** Dielectric slabs completely filling the space between parallel electrodes of a capacitor.*

from one plate to the other, since lines of **D** originate only on *free* charges. Application of Gauss' law to the interface between the two dielectric slabs gives the same result. Then $D_1 = D_2$ or $\epsilon_1 E_1 = \epsilon_2 E_2$ or $K_1 E_1 = K_2 E_2$.

We now find the capacitance from $C = Q_f/V$, where $V = E_1 a + E_2 b$ (obtained from $dV/dx = -E$). Then $C = \sigma_f A/(E_1 a + E_2 b)$. For parallel-plate geometry we have already seen that $D = \sigma_f$ from Gauss' law, so we can write

$$C = \frac{A}{a/\epsilon_1 + b/\epsilon_2} \qquad \text{or} \qquad \frac{A\epsilon_0}{a/K_1 + b/K_2} \qquad \text{farads}$$

5.6 Boundary Conditions at a Dielectric Surface

We now show more rigorously the correctness of the boundary conditions discussed in Sec. 5.4. A plane boundary between regions of

different dielectric constant, K_1 and K_2, is shown in Fig. 5.10. We assume no collection of free charge at the surface. Gauss' law for $\mathbf{D}$ is now applied to a suitable Gaussian surface. The net flux through the volume must be zero since there is no free charge inside the Gaussian surface we have drawn. We can neglect the flux through the curved part of this surface, since this area can be made

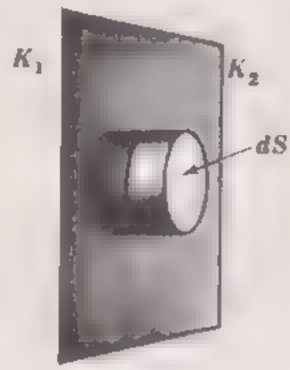


***Fig. 5.10** Gaussian surface for calculation of boundary condition for* $\mathbf{D}$ *across a dielectric boundary. This gives* $D_{n1} = D_{n2}$. *(The normal component of* $\mathbf{D}$ *is continuous across a dielectric boundary.)*

vanishingly small by reducing the distance between the flat surfaces. If we call the flat surface on the left-hand side S_1 and the flat surface on the right S_2, we may write

$$\int_{CS} \mathbf{D}\cdot d\mathbf{S} = \int_{S_1} \mathbf{D}_1\cdot d\mathbf{S} + \int_{S_2} \mathbf{D}_2\cdot d\mathbf{S} = -\int_{S_1} D_{n1}\, dS + \int_{S_2} D_{n2}\, dS = 0$$

The sign for the integral on the left-hand side of the boundary is negative because the lines of $\mathbf{D}$ point *into* the Gaussian volume. It follows directly from the equation above that

$$D_{n1} = D_{n2} \tag{5.20}$$

Thus, *the normal component of* $\mathbf{D}$ *is continuous across the boundary.*

Figure 5.11 shows the argument needed regarding $\mathbf{E}$. Here we invoke the conservation of energy in a static field. We apply the line integral $\oint \mathbf{E}\cdot d\mathbf{l} = 0$ to the path shown. The parts of the path

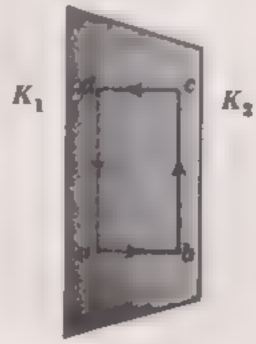


***Fig. 5.11** Line integral for calculation of boundary condition for* $\mathbf{E}$ *across a dielectric boundary. This gives* $E_{t1} = E_{t2}$. *(The tangential component of* $\mathbf{E}$ *is continuous across a dielectric boundary.)*

perpendicular to the plane, ab and cd, can be made vanishingly small so that the integral can be expressed as

$$\int_b^c \mathbf{E}_2 \cdot d\mathbf{l} + \int_d^a \mathbf{E}_1 \cdot d\mathbf{l} = \int_b^c E_{t2}\, dl - \int_d^a E_{t1}\, dl = 0$$

Since $bc = da$,

$$E_{t1} = E_{t2} \tag{5.21}$$

or *the component of* **E** *tangent to the boundary is continuous across the boundary.*

We may use these boundary conditions to show the effect of a dielectric boundary on the direction of an electric field that crosses the boundary. The dielectric materials are assumed isotropic. In such materials **P** is always parallel to **E**, so from Eq. (5.12), **D** is always parallel to **E**. Figure 5.12 shows two vectors corresponding to the directions of the electric field on the two sides of the boundary. We determine the necessary relationship between the angles ϕ_1 and ϕ_2. From the condition $E_{t1} = E_{t2}$ we find $E_1 \sin \phi_1 = E_2 \sin \phi_2$. From the condition $D_{n1} = D_{n2}$ we have $D_1 \cos \phi_1 = D_2 \cos \phi_2$. Combining results for D and E, we get

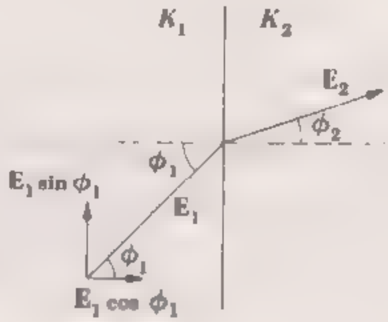


Fig. 5.12 *Change of direction of* **E** *at a dielectric surface.*

$$\frac{E_1 \sin \phi_1}{D_1 \cos \phi_1} = \frac{E_2 \sin \phi_2}{D_2 \cos \phi_2}$$

But since $D_1 = K_1 \epsilon_0 E_1$ and $D_2 = K_2 \epsilon_0 E_2$, our equation becomes

$$\frac{1}{K_1} \tan \phi_1 = \frac{1}{K_2} \tan \phi_2$$

or

$$\frac{\tan \phi_1}{\tan \phi_2} = \frac{K_1}{K_2} \tag{5.22}$$

5.7 *Force between Charges in a Dielectric Medium*

One of the important effects of a dielectric is the modification it produces in the effective force between charges. We consider a spherical charged equipotential body immersed in, say, a dielectric fluid. Let its charge be q_f. The calculation of the force on another charge Q_f at a distance r, due to the charge q_f, proceeds in the usual way: We first calculate the field at Q_f due to q_f, after which the force is readily obtained. We may calculate the field of q_f at a distance r in two ways. Assuming spherical symmetry about q_f, we can evaluate **D** via Gauss' law:

$$\int_{CS} \mathbf{D}\cdot d\mathbf{S} = q_f \qquad \text{or} \qquad D = \frac{q_f}{4\pi r^2} = \epsilon E = \epsilon_0 K E$$

Thus,

$$E = \frac{1}{K}\frac{1}{4\pi\epsilon_0}\frac{q_f}{r^2} = \frac{q_f}{4\pi\epsilon r^2} \tag{5.23}$$

The requirement of spherical symmetry assures that the magnitude of **D** is constant over any spherical surface centered in the charged sphere and that **D** is normal to this surface. This is justified as long as the other charge Q_f is far enough away that the perturbation of the field in the vicinity of q_f is small. We may now take **D** outside the integral, just as we did with **E** when considering the field of a sphere in a vacuum.

Another way to make the calculation of the field due to q_f is to evaluate **E** from Gauss' law, taking into account the polarization charge at the surface of the dielectric where the equipotential sphere begins. We are again assuming spherical symmetry about q_f. Then

$$\int_{CS} \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0}(q_f + q_p)$$

This gives for the field

$$E = \frac{1}{4\pi\epsilon_0}\frac{q_f + q_p}{r^2}$$

Comparing with the earlier result [Eq. (5.23)], we see that the *effective* charge $(q_f + q_p)$ is $(1/K)q_f$.

We next ask for the force due to this field on the other charge Q_f. If we again assume spherical symmetry, this time about Q_f, we

can see that the polarization charge induced in the dielectric surface around the second charge Q_f does not affect the force on it, since, as we have seen, the field inside a spherical distribution of charge due to that charge is zero. We therefore calculate the force from the field in the usual way as

$$F = \frac{1}{4\pi\epsilon}\frac{Q_f q_f}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{Q_f q_f}{Kr^2} \tag{5.24}$$

This is exactly the same as the expression for point charges in vacuum, except that the permittivity of free space ϵ_0 has been replaced by ϵ, the permittivity of the dielectric medium. The requirement of spherical symmetry for the polarization charge density, needed to allow the given solution of Gauss' law, restricts the validity of Eq. (5.24) to situations where the distance between charged bodies is much greater than either radius.

The foregoing calculation gives an example of the general result that in the presence of a dielectric, forces between charges are reduced. This effect is invoked in solid-state physics to explain situations where forces on electrons are decreased by the dielectric nature of the material in which the electrons move. This idea is valid as long as the electron moves far enough in the medium to be subjected to the average field resulting from the polarization.

5.8 Stored Energy in a Dielectric Medium

We have seen that in a vacuum the energy stored in an electric field per unit volume is

$$\tfrac{1}{2}\epsilon_0 E^2$$

We now modify this expression to include the situation in dielectric media. We can proceed just as in Sec. 5.3, accounting for the presence of the dielectric by replacing ϵ_0 by ϵ. This leads us to the expression for energy density,

$$\frac{U}{\text{vol}} = \frac{1}{2}\epsilon E^2 = \frac{ED}{2} = \frac{D^2}{2\epsilon} \qquad \text{joules/m}^3 \tag{5.25}$$

The excess of this stored energy over that in a vacuum is due to the extra work that must be done to polarize the medium.

As in Sec. 4.7, where we calculated the force between capacitor plates by computing the rate of change of stored energy, we shall

now examine a force problem involving a dielectric slab in a capacitor. The problem is to determine the force acting on the slab due to the electric field in the capacitor shown in Fig. 5.13. Before starting

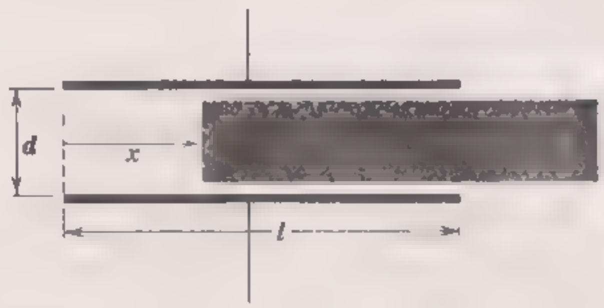


Fig. 5.13 Dielectric slab inserted between plates of a parallel-plate capacitor.

the computation, we observe that the convenient parameter to use in describing the charge state of the capacitor is the potential V or, what amounts to the same thing, E as given by $V = Ed$. This is the same in all regions of the capacitor, whereas the free charge density in the region of the dielectric is different from that in the vacuum region. A decision must be made as to whether the potential or the charge on the capacitor will be held constant. We choose to work the problem as though the capacitor were connected to a source of constant potential such as a battery.

In order to calculate the force on the dielectric slab, we ask for the effect on the stored energy of an elementary displacement dx of the slab. Since this displacement modifies the capacitance of the capacitor, there will be a change dQ in the charge on the capacitor if it is to be held at a constant potential. As a consequence, an amount of work $V\,dQ$ is done by the battery when the slab is moved. Thus the total change in stored energy in the capacitor, dU, is related to the work done by the field in moving the slab by the equation

$$dU = -F\,dx + V\,dQ \tag{5.26}$$

where F is the force exerted by the field on the slab. We proceed to calculate the terms in Eq. (5.26) and solve for the direction and magnitude of F. Applying Eq. (5.25) to the capacitor, we can find the total stored energy. This is

$$\begin{aligned} U &= \tfrac{1}{2}E^2\epsilon_0 xbd + \tfrac{1}{2}E^2\epsilon(l - x)bd \\ &= \tfrac{1}{2}E^2 bd[\epsilon l - (\epsilon - \epsilon_0)x] \end{aligned} \tag{5.27}$$

where b is the width of the capacitor, l the length, and d the plate separation. By differentiating the last equation, we find that the change in U produced by a small displacement dx of the slab is

$$dU = -\tfrac{1}{2}E^2bd(\epsilon - \epsilon_0)\,dx \tag{5.28}$$

The total charge on the capacitor plates is given by

$$Q = \sigma_1A_1 + \sigma_2A_2 = \epsilon_0Exb + \epsilon E(l - x)b$$

where $A_1 = xb$ is the area with no dielectric and $A_2 = (l - x)b$ is the area filled with dielectric. Thus the change in charge for a displacement dx is

$$dQ = -(\epsilon - \epsilon_0)bE\,dx \tag{5.29}$$

Since $V = Ed$, we may now solve Eq. (5.26) for $F\,dx$, obtaining

$$\begin{aligned} F\,dx &= -dU + V\,dQ \\ &= \tfrac{1}{2}E^2bd(\epsilon - \epsilon_0)\,dx - E^2bd(\epsilon - \epsilon_0)\,dx \end{aligned}$$

or

$$F = -\tfrac{1}{2}E^2bd(\epsilon - \epsilon_0) \tag{5.30}$$

Since F is negative, it is in the direction of decreasing x. Hence the electrostatic forces attract the slab *into* the capacitor. The work done by the battery when the slab is displaced is twice the magnitude of the work expended by the field forces in drawing the dielectric into the capacitor. When the plate is allowed to be pulled into the capacitor a given distance, half the energy contributed to the system by the battery is used in doing external work.

Suppose we now make the calculation for the case where the capacitor is held at constant charge instead of being connected to a battery that maintains constant voltage. This would be a more complicated calculation since V would vary as the dielectric slab moved. However, if the slab is in a particular position and the potential has a particular value, the charges on plates and slab are identical whether or not the battery is connected. Since the total force really depends on the forces between free charges on the plates and polarization charges on the dielectric, we must get the same result for the force, whether or not the capacitor is connected to a battery. Thus the force between plates for a given separation and a given potential difference is independent of whether the capacitor is isolated or connected to a constant potential.

5.9 Depolarization Factor

When a polarizable body is placed in an external field, the induced surface charges resulting from the polarization alter the original field. Thus it may be quite difficult to calculate the resultant field inside (or outside) the body. Our earlier discussion of a flat slab placed inside a parallel-plate capacitor was a very special case of this kind of problem. We now discuss some more general cases, avoiding those which are mathematically complicated.

We limit discussion to cases in which the original field $\mathbf{E}_0$ is uniform. We first consider a flat plate of dielectric material placed with its plane perpendicular to the external field (Fig. 5.14). At the

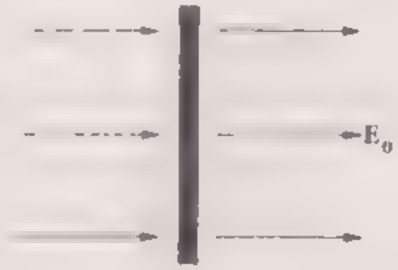


Fig. 5.14 Dielectric plate with plane perpendicular to external field.

edges of the plate the field in and around the plate will be seriously modified and nonuniform as a result of the polarization charges. We shall not try to solve this part of the problem.

In the central region of the plate, however, the polarization charges are uniformly distributed over the surface because the external field remains uniform. We calculate the field inside the plate using the boundary condition already found for $\mathbf{D}$ The result was that $D_{n1} = D_{n2}$; that is, the normal components of $\mathbf{D}$ are the same inside and outside the boundary. As long as we stay far away from the edges of the plate, we can obtain $\mathbf{D}$ outside the plate from $\mathbf{D}_0 = \epsilon_0\mathbf{E}_0$. From the boundary condition we then have

$$\mathbf{D}_{\text{plate}} = \mathbf{D}_0 = \epsilon_0\mathbf{E}_0$$

On the other hand, the electric field inside the plate, $\mathbf{E}_{\text{plate}}$, is related to $\mathbf{D}_{\text{plate}}$ by the equation $\mathbf{D}_{\text{plate}} = \epsilon\mathbf{E}_{\text{plate}}$. Combining these results, we find

$$\mathbf{E}_{\text{plate}} = \frac{\epsilon_0}{\epsilon}\mathbf{E}_0 = \frac{1}{K}\mathbf{E}_0 \tag{5.31}$$

That is, the field inside the plate is reduced by a factor $\dfrac{\epsilon_0}{\epsilon} = \dfrac{1}{K}$ relative to the original field outside.

It is convenient to look upon the field in the plate as resulting from the original external field E_0, modified by the uniform *depolarizing field* of the polarization charge. Thus we may write

$$E_{\text{plate}} = E_0 - E_{dep} \tag{5.32}$$

where $\mathbf{E}_{dep}$ is the depolarizing field of the polarization charges on the surfaces of the plate. This field acts in opposition to the external field E_0. Combining Eqs. (5.31) and (5.32), we solve for the depolarizing field,

$$E_{dep} = (K - 1)E_{\text{plate}} \tag{5.33}$$

Now the polarization of the dielectric is related to the field in the dielectric by the equation

$$P = \epsilon_0 \chi E_{\text{plate}} = \epsilon_0 (K - 1)E_{\text{plate}}$$

Comparison with Eq. (5.33) gives

$$E_{dep} = \frac{P}{\epsilon_0}$$

This case of a flat plate is a special one. In general, the depolarization field is less than for this geometry. For the general case, we write

$$E_{dep} = \frac{1}{\epsilon_0} LP \tag{5.34}$$

where L is called the *depolarizing factor*, a quantity that depends on the shape of the dielectric. The factor L gives a measure of how much the original external field E_0 is modified for different sample shapes.

We can evaluate L for the flat-plate geometry by remembering that $P = \epsilon_0 \chi E_{\text{plate}} = \epsilon_0(K - 1)E_{\text{plate}}$. Combining this with Eqs. (5.33) and (5.34), we find $L = 1$. This is the maximum value L can have, so we have chosen the geometry giving the maximum effect from polarization charges induced on the surface.

The case giving the opposite extreme (minimum value of L) is that of a long, thin rod placed with its axis parallel to the field

(Fig. 5.15). The appropriate boundary condition in the central region of the rod, away from the perturbing effects of the polarization charge induced at the ends, is $E_{t1} = E_{t2}$. For this shape, the field

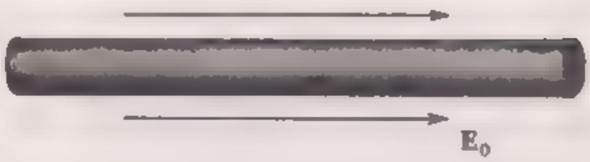


Fig. 5.15 Thin dielectric rod with axis parallel to external field.

inside and outside is tangential to the surface, so we have, directly,

$$E_{\text{rod}} = E_0$$

Since E_{rod} is equal to the unmodified external field, the depolarizing field in this situation must be zero, so we have

$$L = 0$$

All values of L lie between these two extremes of 1 and 0.

Strictly speaking, we can discuss the value of the depolarizing factor only if the depolarizing field is uniform, since only in such cases will the field inside a body placed in a uniform field be uniform. In the two examples of the plate perpendicular to the field and the rod placed parallel to the field, the approximation of uniform field is good only in the central regions. For all general ellipsoids in uniform applied fields the depolarizing field is uniform, giving uniform fields inside the bodies when the applied field is uniform. Thus if we distort the flat plate into a flat ellipsoid and distort the rod into a long thin ellipsoid, we obtain cases for which the field is constant throughout the entire body. Another shape of interest is the sphere, which is a special case of the general ellipsoid. The field inside a dielectric sphere is uniform, and the depolarizing factor is $L = \frac{1}{3}$. This value is obtained in Appendix C. Exactly similar problems arise for magnetic bodies immersed in uniform magnetic fields, as we discuss in Chap. 9.

5.10 Atomic Polarizability

Of all the parameters of dielectrics studied, the susceptibility χ comes the closest to telling about the polarizability of the atoms in matter. Thus in the equation $\chi = P/\epsilon_0 E$, the dipole moment per unit volume is related to the macroscopic field in the material. The trouble is, however, that the macroscopic or average field is

not a satisfactory measure of the *local field* producing the polarization of each atom. Let us call this local field E_{loc} and postpone its evaluation for the moment. We can then define the *atomic polarizability* α by

$$p = \alpha E_{loc} \tag{5.35}$$

where p is the induced atomic dipole moment produced by the local field. We then write for the polarization

$$P = Np = N\alpha E_{loc} \tag{5.36}$$

where N is the number of atoms per unit volume and we have assumed only one kind of atom in the dielectric material. Here we do not attempt to evaluate α from atomic properties, but we do indicate the problems involved in evaluating E_{loc}.

Let us take an ellipsoidal sample of dielectric placed in a uniform field E_0 as shown in Fig. 5.16. We center attention on

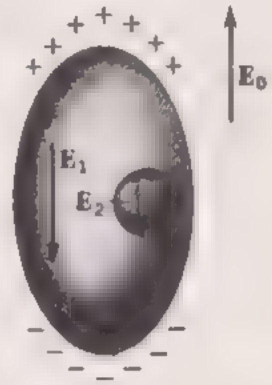


***Fig. 5.16** Calculation of E_{loc} at a point P within an ellipsoidal dielectric placed in a uniform electric field. All dipoles outside the spherical boundary have effects that can be replaced by effects of induced surface charges. Charges within the spherical boundary must be treated as individual dipoles. Their effects cancel for crystals having cubic symmetry.*

some particular atom in the solid and inquire about the terms contributing to E_{loc} at the atomic site (excluding, of course, the atom itself, since it will not be polarized by its own field). The field at the atom is the sum of the external field plus contributions from all the other atoms (dipoles) in the material. We have already seen that the contribution from atoms *far away* from our atom can be computed from the equivalent effects of surface polarization charge. We shall discuss contributions from dipoles close to this atom separately. We draw a spherical boundary around the atom, of such size that we may safely consider all atoms outside the boundary to be far away. Their effects are accounted for on the

basis of surface polarization charges. Excluding contributions from nearby atoms (inside the imaginary boundary), the local field is due to three sources,

$$\mathbf{E}_{loc} = \mathbf{E}_0 + \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3$$

$\mathbf{E}_0$ is the external applied field, and $\mathbf{E}_1$ the depolarization field due to the polarization charge on the outer surface of the dielectric body. The magnitude of $\mathbf{E}_1$ depends, of course, on the shape of the sample—in particular on the value of L, the depolarization factor—and is given by $-L\mathbf{P}/\epsilon_0$. The term $\mathbf{E}_2$ similarly results from the induced polarization charge on the inner surface of the imaginary sphere and is given by $\frac{1}{3}(\mathbf{P}/\epsilon_0)$. The depolarizing factor $\frac{1}{3}$ is the same one we found for the dielectric sphere, but because it is an inside surface inside a dielectric, its contribution is positive here. It is *not* the same as we would compute for a spherical hole inside a dielectric since we do not imagine that the atoms are removed from the excluded sphere. Thus in our calculation, the average field is uniform, whereas a real spherical hole gives a nonuniform average field in the region outside the hole.

We are now left with the task of evaluating $\mathbf{E}_3$, the local contribution of the set of dipoles in the excluded sphere. It turns out, however, that if we take the rather common case of a lattice having cubic symmetry, the contributions of these nearby dipoles cancel out. Therefore, we may write for the local field

$$\mathbf{E}_{loc} = \mathbf{E}_0 - \frac{L\mathbf{P}}{\epsilon_0} + \frac{1}{3}\frac{\mathbf{P}}{\epsilon_0}$$

For a more complicated crystal structure, a further term may be required to account for the nearby dipoles. We need not take into account separately the polarization charges on the outer surface of the excluded sphere since these are already included in the contribution from the nearby dipoles.

Once we are able to relate E_{loc} to the applied field, as above, we are in a position to determine the atomic polarizability from experiment. We shall say only a few words about its theoretical estimation. There are three kinds of possible contributions to the polarizability. In all situations the application of an electric field causes some distortion of the atomic electronic cloud relative to the nucleus. We may call this the *electronic polarizability*. Second, if we

have a solid made up of ions, there is relative motion of positive and negative ions to produce *ionic polarizability*. Finally, if there are molecules with permanent dipole moments involved, as for example, with water molecules, an external field tends to orient the otherwise randomly oriented dipoles so as to produce a net addition to the polarization. This last is called the *dipolar polarizability*.

These three types of response to electric fields may be somewhat sorted out by experiments in which the applied field is rapidly varied. As the frequency of the applied field is gradually increased, at first all three types of polarizability contribute. When frequencies up in the microwave range (say 10^9 cps) are reached, the dipoles with their relatively slow response are not able to follow the rapidly varying electric field and their contribution to polarization ceases. Ionic polarization continues to act up to the infrared frequency range (say 10^{13} cps), and electronic polarization continues up to the ultraviolet frequency range (say 10^{15} cps).

Polarization from permanent dipoles is important in all liquids and gases whose molecules have permanent dipoles, and it also occurs in some solids. It is the damping out of this kind of response that accounts for the decrease in the dielectric constant of water from 81 for static- or low-frequency fields to 1.77 at optical frequencies. Since thermal vibrations tend to upset the orientation of permanent dipoles by an electric field, dipolar polarization decreases with increasing temperature.

5.11 Ferroelectric Materials

Certain types of dielectric materials become spontaneously polarized in the absence of external fields. This self-polarization results from the displacement of ions due to local electric fields. The local fields set up by ion displacement produce forces on the ions that are greater for small displacements than the elastic restoring forces within the crystal. As a result, the equilibrium position of the ions is such as to give a net polarization in the crystal. Such materials are called *ferroelectric* by analogy with the somewhat similar magnetic effects in ferromagnetic materials (see Sec. 9.10). Ferroelectric materials often exhibit large susceptibility. They show hysteresis in the polarization produced by external fields and lose their spontaneous polarization above some critical temperature. One of the best-known ferroelectric materials is barium titanate ($BaTiO_3$),

which, because of its high dielectric constant, is often used in capacitors, especially where small size and large capacitance are required.

PROBLEMS

5.1 A dielectric block such as shown in Fig. P5.1 is uniformly polarized. The polarization is **P**. Find the polarization charge density σ_p on the faces 1, 2, and 3. (Find both magnitude and sign of the charge.)

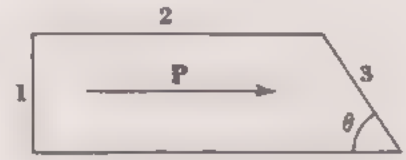


Fig. P5.1

5.2 Two similar dielectric ellipsoids are placed in an electric field as shown in Fig. P5.2. For which orientation is the depolarization factor larger? Give qualitative reasons.

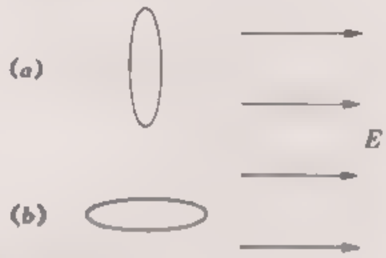


Fig. P5.2

5.3 The voltage between parallel plates of a capacitor is V_1. The plates are isolated electrically. A dielectric slab of dielectric constant K is inserted between the plates and completely fills the volume between them. Find the new potential V_2. Compare the stored energy before and after inserting the slab. On the basis of this comparison, make an argument as to whether electrostatic forces pull the slab into the space between plates or tend to push it away.

5.4 Suppose the capacitor in Prob. 5.3 were connected to a battery so as to maintain constant voltage when the dielectric slab is inserted between the plates. Compare the stored energy in the capacitor before and after inserting the dielectric. On the basis of this comparison, can you argue about the direction of the force on the slab, as in Prob. 5.3? Explain.

5.5 At a boundary between two dielectric materials having different dielectric constants, show by means of a simple sketch that although *lines* of **D** are continuous, the *magnitude* of **D** is generally different in the two materials. Show by means of another sketch the special condition under which **D** has the same value in both materials.

5.6 Compare the capacitances of two identical capacitors with dielectrics inserted as shown in Fig. P5.6 and having dielectric constants K_1 and K_2.

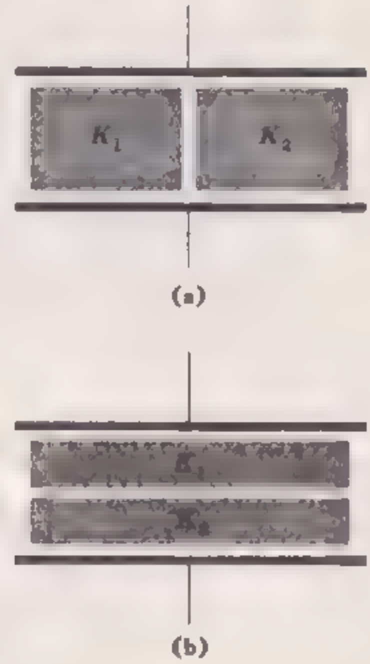


Fig. P5.6

5.7 Discuss a metal as a polarizable body. What is the value of the polarization **P**? What is the susceptibility χ?

5.8 A sphere of radius r is polarized uniformly and has a polarization **P** in the x direction. Write the expression for the surface polarization charge for a ring whose radius vector makes an angle θ with the x axis as shown in Fig. P5.8. Integrate this expression to get the total positive surface polarization charge on the sphere.

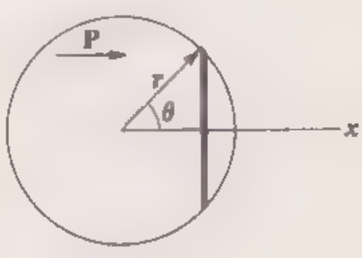


Fig. P5.8

5.9 Two capacitors of equal capacitance C are connected in parallel, charged to a voltage V_1, and then isolated from the voltage source, as shown in Fig. P5.9. A dielectric of dielectric constant K is inserted

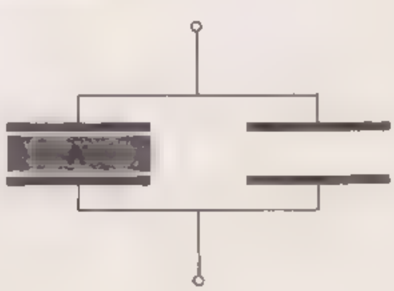


Fig. P5.9

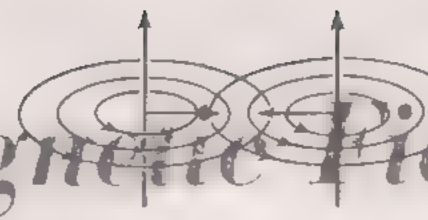
into one capacitor and completely fills the space between plates. Calculate the free charge transferred from one capacitor to the other and the final voltage V_2 across the capacitors in terms of C, V_1, and K.

5.10 The capacitance of a capacitor is increased by a factor of 1.5 when it is completely filled with a certain dielectric material. Find the dielectric constant of the material and its electric susceptibility.

5.11 The angle of incidence of the electric field at a plane dielectric boundary is 20°. Find the angle of refraction within the medium if the dielectric constant of the medium is 1.25. Assume vacuum outside the medium.

SIX

The Magnetic Field



6.1 Introduction

The first five chapters have dealt with the effects of charges at rest. This study of electrostatics forms a very important part of our entire work, but, as we shall see, many entirely new phenomena occur when we deal with charges in motion. Electromagnetic theory provides a single framework for discussing the seemingly widely divergent phenomena of electricity and magnetism. Later we shall discuss the unifying ideas of the theory. First we examine the effects of moving charges, called *magnetism*.

Historically, magnetism began with the study of interactions between ferromagnetic materials; substances such as iron under appropriate conditions exhibit strong forces of attraction and repulsion, which resemble, but are quite distinct from, electrostatic forces. An early example of the knowledge of magnetism was the use of a permanent magnet in the earth's magnetic field, as a compass for navigational purposes. In 1819 Oersted first showed a connection between electricity and magnetism by demonstrating the

torque on a compass needle caused by a nearby electric current. We shall delay the discussion of magnetic materials until later and, instead, first discuss magnetism in terms of the forces between the moving charges of current elements. Students who are not already familiar with the concepts of current and current density, which we shall be using, are referred to the early part of Chap. 7.

6.2 The Magnetic Force between Current Elements

We introduce the study of magnetism by inquiring about the force between two parallel wires in a vacuum, carrying currents i_1 and i_2 (Fig. 6.1). The experimental fact, as first studied by Ampère, is that

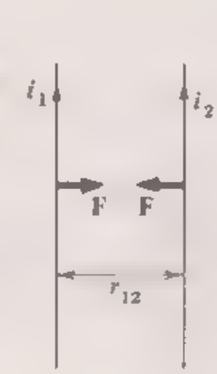


***Fig. 6.1** Two parallel wires carrying current in the same sense, showing forces of attraction.*

the two wires are attracted to each other. If we reverse one current, the force becomes one of repulsion. The forces are modified if either the current magnitudes, the shape, or the relative positions of the conductors are changed.

There is a difficulty in discussing this problem. Any current element must of necessity be a part of a complete loop or circuit. Thus a realistic formulation must be in terms of the force between two complete loops. We may, however, consider the simplified case as long as we realize that comparison with experiment involves accounting for the forces on entire circuits.

There is no possibility of confusing this force between currents with electrostatic forces. In the first place, there is in general no net charge on the conductors. The charge density of conduction electrons just compensates the positive charge on the lattice ions. Second, the force is reversed in sign by reversing the *direction* of either current. The *magnetic* force is thus associated with *moving* charges.

We quote briefly from the original paper of Ampère [1] in which he describes the conclusions to be drawn from his experiments on the forces between currents:

[1] André Marie Ampère, Mutual Interactions between Two Electric Currents, *Ann. chim. et phys.*, (II) 15:59–76 (1820).

I have discovered some . . . remarkable . . . [effects] . . . by arranging in parallel directions two straight parts of two conducting wires joining the ends of two voltaic piles; the one was fixed and the other, suspended on points and made very sensitive to motion by a counterweight, could approach the first or move from it while keeping parallel with it. I then observed that when I passed a current of electricity in both of these wires at once they attracted each other when the two currents were in the same sense and repelled each other when they were in opposite senses. Now these attractions or repulsions of electric currents differ essentially from those that electricity produces in a state of repose: first, they cease, as chemical decompositions do, as soon as we break the circuit of the conducting bodies; secondly, in the ordinary electric attractions and repulsions the electricities of opposite sort attract and those of the same name repel, in the attractions and repulsions of electric currents we have precisely the contrary . . . . Thirdly, in the case of attraction, when it is sufficiently strong to bring the movable conductor into contact with the fixed conductor, they remain attached to one another like two magnets and do not separate after a while, as happens when two conducting bodies which attract each other because they are electrified, one positively and the other negatively, come to touch. Finally, and it appears that this last circumstance depends on the same cause as the preceding, two electric currents attract or repel in vacuum as in air, which is contrary to that which we observe in the mutual action of two conducting bodies charged with ordinary electricity. It is not the place here to explain these new phenomena; the attractions and repulsions which occur between two parallel currents, according as they are directed in the same sense or in opposite senses, are facts given by an experiment which it is easy to repeat.

The experimental fact is that for a current element $i_1\, dl_1$ (a current i_1 flowing through a length dl_1) parallel to another element

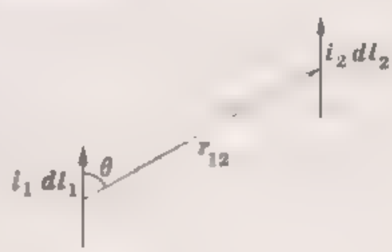


***Fig. 6.2** The force between two parallel current elements $i_1\, dl_1$ and $i_2\, dl_2$ as given by Eq. (6.1).*

$i_2\, dl_2$ and separated from it by a distance r_{12} as shown in Fig. 6.2, the mutual force is given by

$$dF = \frac{\mu_0}{4\pi} \frac{1}{{r_{12}}^2}\, i_1\, dl\, i_2\, dl_2 \sin\theta \tag{6.1}$$

The magnitude of the force between current elements at a given distance apart is a fundamental characteristic of nature. The constant of proportionality is $\mu_0/4\pi$; the numerical value of this constant depends on the units of measurement used for current, distance, and force. In the mks system, using amperes, meters, and newtons,

$$\mu_0 = 4\pi \times 10^{-7} \qquad \text{webers/amp-m}$$

This is called the *permeability* of free space. Equation (6.1) refers to a very special case, that of parallel current elements. If we had to deal only with parallel current elements, this equation would be the only one needed. The more complicated expression shortly to be described is necessary in order to explain the changes in the forces when there is an arbitrary angle between the current elements. This expression involves no mention of a magnetic field. In both electrostatics and magnetism, fields are invoked to simplify the reasoning required. Equation (6.1) cannot be used directly since it is impossible in practice to set up isolated current elements. To be of practical use, this equation would have to be integrated over the entire current path for both i_1 and i_2.

We now reformulate the problem in terms of a field. If the mathematics seems a little more complicated than in the case of electrostatics, this is only because the physical situation is itself more complex. Our procedure is to describe the *magnetic induction field* (a vector quantity $\mathbf{B}$) due to a current element and then to state the force on another current element placed in that field. This gives the general result we want, of which Eq. (6.1) is a special case.

Experiments on the force between two current circuits, when formulated in terms of the magnetic induction field, show that the experimental results can be explained if each small current element $i\,d\mathbf{l}$ gives rise to a contribution to the total magnetic induction field according to the equation

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\,\frac{i\,d\mathbf{l} \times \hat{\mathbf{r}}}{r^2} \tag{6.2}$$

where $\hat{\mathbf{r}}$ is again a unit vector, having a magnitude of unity and the direction of the line drawn *from* the current element *to* the point at

which $d\mathbf{B}$ is being determined. This is the law of Biot and Savart. In the mks system the unit of $\mathbf{B}$ is the weber per square meter. This unit is discussed later in the chapter. We see from Eq. (6.2) that the units of permeability μ_0 are webers per ampere-meter, as given above. Since Eq. (6.2) gives our first use of the *vector cross product*, we now pause briefly to review the mathematics involved.

The vector cross product The special feature of Eq. (6.2) is that it expresses a rather complicated relationship between the directions and magnitudes of the vectors $d\mathbf{B}$, $d\mathbf{l}$, and $\mathbf{r}$. Fortunately, there is a well-known mathematical expression that exactly describes the situation. We explain this in terms of three vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, which are related by the equation $\mathbf{A} = \mathbf{B} \times \mathbf{C}$, where $\mathbf{A}$ is called the vector cross product of $\mathbf{B}$ and $\mathbf{C}$.

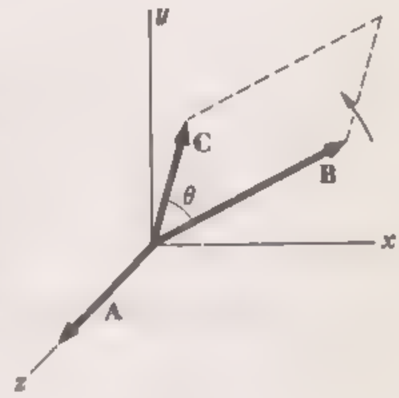


Fig. 6.3 ***The vector*** $\mathbf{A}$ ***is the vector cross product of vectors*** $\mathbf{B}$ ***and*** $\mathbf{C}$***:*** $\mathbf{A} = \mathbf{B} \times \mathbf{C} = BC \sin\theta$. $\mathbf{A}$ ***is perpendicular to the plane of*** $\mathbf{B}$ ***and*** $\mathbf{C}$***, and its magnitude is equal to the area of the parallelogram built on*** $\mathbf{B}$ ***and*** $\mathbf{C}$***.***

In Fig. 6.3 we show vectors $\mathbf{B}$ and $\mathbf{C}$ placed in the xy plane of a right-hand coordinate system. The resultant vector $\mathbf{A}$ of the cross product $\mathbf{B} \times \mathbf{C}$ is perpendicular to both $\mathbf{B}$ and $\mathbf{C}$ and is therefore in the z direction. The sign of $\mathbf{A}$ is obtained by imagining the rotation of the vector given by the first term in the cross product, $\mathbf{B}$, about its tail in the direction of the curved fingers of the right hand toward the second vector $\mathbf{C}$, and taking for the direction of the resultant vector $\mathbf{A}$ the direction of the extended thumb. This gives the result shown in the figure. We see that $\mathbf{B} \times \mathbf{C} = -(\mathbf{C} \times \mathbf{B})$, since in the latter case $\mathbf{C}$ is rotated into $\mathbf{B}$ and the thumb will point along the negative z axis.

The magnitude of the resultant vector $\mathbf{A}$ is given by $BC \sin\theta$, where θ is the angle between the two vectors that form the cross product. This is easily remembered by noting that $BC \sin\theta$ is the magnitude of one vector times the component of the other vector *perpendicular* to the first vector. This quantity is the area of the parallelogram constructed using the two vectors as adjacent sides as shown in Fig. 6.3. Thus the magnitude of the cross product of two vectors that are at right angles is the product of the magnitudes of the two vectors. The cross product of two parallel vectors is zero.

We can now interpret Eq. (6.2) on the basis of the meaning of the cross product. We first write Eq. (6.2) in an alternative form,

$$dB = \frac{\mu_0}{4\pi}\frac{i\,d\mathbf{l} \times \mathbf{r}}{r^3} \tag{6.2a}$$

This has the same meaning as Eq. (6.2) and is obtained from it by using the relationship between $\hat{\mathbf{r}}$ and $\mathbf{r}$,

$$\hat{\mathbf{r}} = \frac{\mathbf{r}}{r}$$

In Fig. 6.4 the vectors $i\,d\mathbf{l}$ and $\mathbf{r}$ are drawn in the plane of the paper.

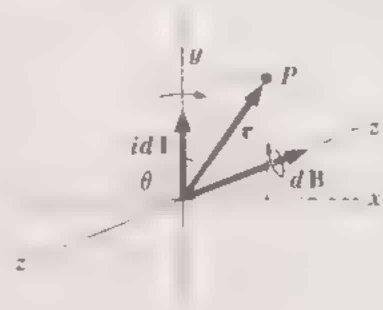


***Fig. 6.4** **dB** is the contribution to the magnetic induction field at P of the current element i d**l**. **dB** is perpendicular to both i d**l** and **r** and is directed into the paper.*

The resultant vector $d\mathbf{B}$, according to Eq. (6.2) or (6.2a), has a magnitude

$$dB = \frac{\mu_0}{4\pi r^3}\,i\,dl\,r\sin\theta = \frac{\mu_0}{4\pi r^2}\,i\,dl\sin\theta$$

and is pointed into the paper along the $-z$ direction. This is the value of $d\mathbf{B}$ at the point P, not at the point where the tails of the two vectors meet. This factor is implicit in the equation in which $\mathbf{r}$ is the vector *from* the current element *to* the point in space where we are finding $d\mathbf{B}$. As we are writing Eq. (6.2), the r^2 in the denominator is the magnitude of this distance and $\hat{\mathbf{r}}$ is the unit vector giving the direction of $\mathbf{r}$. In effect, $\hat{\mathbf{r}}$ is used only to define the value of θ. It is more usual to draw the vectors as shown in Fig. 6.5. This

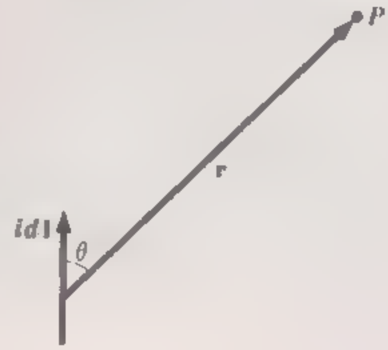


***Fig. 6.5** Vectors i d**l** and **r** as used for obtaining field contribution **dB** at point P. **dB** at the point P is perpendicular to both i d**l** and **r** and is directed into the paper.*

is completely equivalent to Fig. 6.4, since $i\,d\mathbf{l}$ is of infinitesimal length $d\mathbf{B}$ at the point P is perpendicular to both $i\,d\mathbf{l}$ and $\mathbf{r}$ and is directed into the paper.

The fact that we have, as in electrostatics, an inverse-square law leads again to the usefulness of the concept of lines of force. As with electric field lines, lines of magnetic induction are continuous in space and give field direction by their direction, and magnitude by their density. However, in one respect there is a very great difference between electric and magnetic field configurations. Magnetic lines have no sources (as do electric field lines, on electric charges) but are continuous and join back on themselves. An examination of Eq. (6.2) and Fig. 6.5 shows this to be true. Let the point P move around the current axis at a constant distance from the axis. From Eq. (6.2), the magnetic induction vector $d\mathbf{B}$ is constant along this path and at each point has a direction tangent to the path. These are just the requirements for the lines to be concentric circles around the current. Figure 6.6 shows some lines

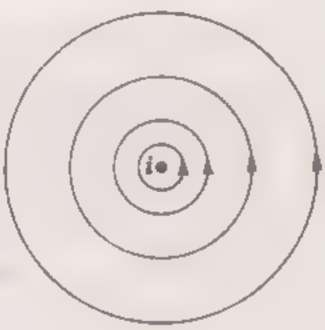


***Fig. 6.6** Lines of magnetic induction around a long wire carrying current upward out of the paper.*

around a current element. If the right-hand thumb points in the direction of (positive) current, the lines are pointed in the direction of the fingers curled around the current element. This *right-hand rule* is an expression of the result of the cross product given in Eq. (6.2).

Before integrating Eq. (6.2) to obtain the field configuration around some extended current elements, we shall set up the other rule that must be employed if we are to use the magnetic induction field $\mathbf{B}$ in calculating the force between circuit elements. The force on a current element $i\,d\mathbf{l}$, as inferred from experiments on closed current loops, is

$$d\mathbf{F} = i\,d\mathbf{l} \times \mathbf{B} \tag{6.3}$$

This means $i\,dl\,B \sin\theta$, in a direction given by $d\mathbf{l} \times \mathbf{B}$ (perpen-

dicular to the plane of $d\mathbf{l}$ and $\mathbf{B}$), again in the sense of the right-hand rule.

We apply these two equations to the two parallel current elements sketched in Fig. 6.1. The first step in calculating the force between the two current elements is to calculate the magnetic induction field $d\mathbf{B}$ at, say, the right-hand conductor due to the current in the left-hand conductor. This is given by Eq. (6.2) as

$$dB = \frac{\mu_0}{4\pi} \frac{|i_1\, d\mathbf{l}_1 \times \hat{\mathbf{r}}|}{r^2} = \frac{\mu_0}{4\pi r^2} i_1\, dl_1 \tag{6.4}$$

for this special case of $i_1\, d\mathbf{l}_1 \perp \mathbf{r}$. The direction of $d\mathbf{B}$ is perpendicular to the plane of the paper ($\perp$ to $d\mathbf{l}$ and $\mathbf{r}$), and the cross-product rule puts it *into* the paper in this case. (The vector $\mathbf{r}$ points *from* the current element causing the field.)

Next, the force equation (6.3) is applied to the right-hand conductor, giving

$$dF = |i_2\, d\mathbf{l}_2 \times d\mathbf{B}| = i_2\, dl_2\, dB \tag{6.5}$$

for $i_2\, d\mathbf{l}_2 \perp d\mathbf{B}$. Combining Eqs. (6.4) and (6.5), we find exactly the expression written in Eq. (6.1), for the case where θ, the angle between $i\, d\mathbf{l}$ and $\mathbf{r}$, is 90°. In this simple example it may seem that the introduction of $\mathbf{B}$ was unnecessary. Later more complicated examples will fully justify its use.

From the expressions for $\mathbf{B}$ and $\mathbf{F}$ in their vector form, it is clear that the correct general expression for the force between two current elements is given by

$$d\mathbf{F} = \frac{\mu_0}{4\pi} \frac{i_2\, d\mathbf{l}_2 \times (i_1\, d\mathbf{l}_1 \times \hat{\mathbf{r}}_{12})}{{r_{12}}^2} \tag{6.6}$$

where $\hat{\mathbf{r}}_{12}$ is the vector pointing *from* the first *to* the second current element. The complexity of this expression, involving two cross products, is to be blamed on the complexity of nature itself.

The result of Eq. (6.6) looks as though it violates Newton's third law. Thus, suppose the two current elements are in the plane of the paper but not parallel, as shown in Fig. 6.7. $d\mathbf{B}$ at (2) will

***Fig. 6.7** Force between two current elements according to Eq. (6.6).*

be into the paper as before, but the force on the second element will have an upward component toward the top of the page, as will the force on the first element. Thus there is a net force upward from this pair of current elements. This difficulty is explained by the fact that actually Eq. (6.6) is complete only when we integrate the expression over the entire current loop of which $i_1\,dl_1$ and $i_2\,dl_2$ are only differential elements. It can be shown that the net force on the whole system is then always zero. From the experimental point of view, there is no difficulty since it is impossible to set up an isolated current element. Whenever Eqs. (6.2) and (6.3) are used for determining forces between complete circuits, integration over these circuits gives no net force on the whole system. Another seemingly troublesome case is that of current elements resulting from isolated charges moving with a velocity v, rather than from the flow of continuous charge. Here again, as we shall see later, there is no real difficulty.

6.3 Examples

In this section we give some examples of the calculation of the magnetic induction **B** for some simple arrangements of conductors.

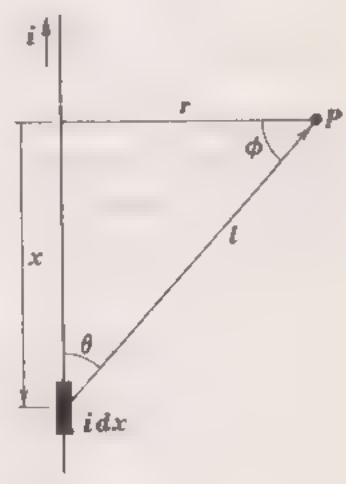


***Fig. 6.8** Magnetic induction due to current in a straight wire.*

a Calculation of **B** *for points out from a long straight wire carrying a current i* A point P is chosen at a fixed distance r from the wire (see Fig. 6.8). A representative current element $i\,dx$ is at a distance l from the chosen point P. Applying Eq. (6.2), we find

$$dB_p = \frac{\mu_0}{4\pi}\,\frac{i\,dx \sin\theta}{l^2}$$

where $d\mathbf{B}_p$ is directed into the paper. Before integrating this, we must express the variables l and θ in terms of a single variable. It is convenient to choose ϕ as the variable and to write

$$B = \frac{\mu_0 i}{4\pi} \int_{-\pi/2}^{\pi/2} \frac{dx}{l^2} \cos\phi$$

The vector addition of the contributions to $\mathbf{B}$ from different current elements is simply arithmetical since all contributions $d\mathbf{B}$ point in the same direction. The integration covers from $-\pi/2$ to $+\pi/2$ to allow for an infinitely long wire. To replace x and dx, we use the relationships $x = r\tan\phi$, $dx = r\sec^2\phi\, d\phi$, $l = r\cos\phi$. Then

$$B = \frac{\mu_0 i}{4\pi r} \int_{-\pi/2}^{\pi/2} \cos\phi\, d\phi = \frac{\mu_0 i}{2\pi r} \qquad \text{webers/m}^2 \tag{6.7}$$

Thus the lines of $\mathbf{B}$ form concentric circles around the wire. The field falls off as $1/r$.

b Calculation of the magnetic induction along the axis of a current loop We choose a point P along the axis of the loop (see Fig. 6.9) and again evaluate

$$d\mathbf{B} = \frac{\mu_0}{4\pi} \frac{i\, d\mathbf{l} \times \hat{\mathbf{r}}}{r^2} \tag{6.2}$$

Since, for all elements around the loop, $\mathbf{r} \perp i\, d\mathbf{l}$, the value of $\sin\theta$ in the cross product here is 1. However, in taking the vector

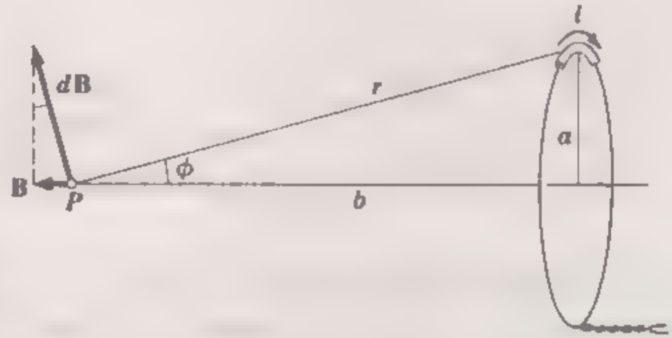


Fig. 6.9 Magnetic induction along the axis of a current loop.

sum of the vectors $d\mathbf{B}$ we must take account of the fact that each $d\mathbf{B}$ is perpendicular to $\mathbf{r}$ at the point P. Because of symmetry, all components of $\mathbf{B}$ not parallel to the axis b will cancel when the equation is integrated around the loop. Therefore the resultant

field will be given by summing only the components $d\mathbf{B} \sin\theta$ along the axis. Thus B is given by

$$B = \frac{\mu_0 i}{4\pi} \frac{\sin\phi}{r^2} \int dl$$

Sin ϕ and r are constant, so they are taken outside the integral. The length dl integrated around the loop is $2\pi a$, so if $\sin\phi$ and r are expressed in terms of the constants a and b, we find

$$B = \frac{\mu_0 i}{2} \frac{a^2}{(a^2 + b^2)^{3/2}} \qquad \text{webers/m}^2 \tag{6.8}$$

At the center of the loop $b = 0$ and $B = \mu_0 i/2a$, pointing along the axis of the loop in a direction depending on the direction of current flow. If the coil has N closely packed turns, the result is multiplied by N. Although we have not accounted for the current leads to the loop, their contribution to the field can be made vanishingly small by keeping them close together. Since the lines of magnetic induction are concentric circles around the wires in opposite directions for the two wires, to a good approximation they cancel when close together as in the lead wires.

When we choose the point P far from the loop ($b \gg a$), the equation for B may be written

$$B = \frac{\mu_0}{4\pi} \frac{2iA}{r^3} \qquad \text{webers/m}^2 \tag{6.9}$$

Here we have written $A = \pi a^2$, the area of the loop. The quantity iA is called the *magnetic dipole moment* for reasons developed later, when we discuss in more detail the field around a current loop. We shall show that $\mathbf{B}$ at large distances from the loop depends only on the area and not on the shape of the loop.

The direction of $\mathbf{B}$ depends on the direction of the current in the loop. We can see from Fig. 6.9 that when we look at the current loop from the left, counterclockwise current gives a magnetic induction field directed toward the viewer. We adopt the sign convention that B is positive when it is directed *outward* from a current loop toward the viewer, and the current is positive when it is directed *counterclockwise* as we view it. This convention allows an unambiguous interpretation of equations like (6.9).

c Calculation of the field along the axis of a solenoid A coil wound as a spiral on a cylinder is called a *solenoid.*[1] In order to find the field at a point P inside a solenoid (on its axis, as shown in Fig. 6.10), we use the results of the preceding problem for the field along the axis of a loop. Since we need to know the number of turns per unit

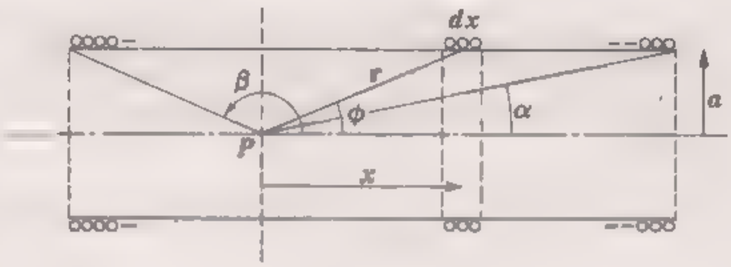


Fig. 6.10 Magnetic induction inside a solenoid.

length, we take N for the total turns and L for the length of the solenoid. We then apply Eq. (6.8) to a thin section of the solenoid of width dx with a total current $i(N/L)\,dx$:

$$dB = \frac{\mu_0 Ni}{2L} \frac{a^2}{(a^2 + x^2)^{3/2}}\,dx$$

This equation must now be integrated, and we choose ϕ for the variable. Making the substitutions $x = a \cot \phi$ and $dx = -a \csc^2\phi\, d\phi$, the equation becomes

$$B = -\frac{\mu_0 Ni}{2L}\int_\beta^\alpha \sin \phi\, d\phi = \frac{\mu_0 Ni}{2L}(\cos \alpha - \cos \beta)$$

If we have chosen the point P in the middle of a long solenoid, $\alpha = 0$ and $\beta = 180°$, and we have

$$B = \frac{\mu_0 Ni}{L} \qquad \text{webers/m}^2 \tag{6.10}$$

Choosing P at one end of the solenoid, we have $\alpha = 90°$, $\beta = 180°$, and

$$B = \frac{\mu_0 Ni}{2L} \qquad \text{webers/m}^2 \tag{6.11}$$

Thus the field strength at the end of a long solenoid is just one-half that at the center.

[1] It was Ampère who first thought of obtaining more intense magnetic fields by passing current through a wire wound in a closely spaced spiral He thought of the cylindrical coil as a "canal" in which lines of magnetic field are confined Hence his use of *solenoid* from the Greek word for channel.

Looking again at Eq. (6.10), which gives the magnetic induction field inside a long solenoid, we shall call the quantity Ni/L the *solenoidal current density*. That is, it is the number of amperes circulating around the coil per meter of length of coil. The result is the same if we double the number of turns per meter and reduce the current to half the original value or if we make any other change that keeps Ni/L constant. We may thus write

$$\frac{Ni}{L} = j^s \qquad \text{amp/m or amp-turns/m}$$

where j^s is the solenoidal current density. We may then write for Eq. (6.10)

$$B = \mu_0 j^s \qquad \text{webers/m}^2 \tag{6.12}$$

where j^s is the solenoidal current density. This quantity is to be distinguished from the conventional current density j, which is defined in Sec. 7.2 as the volume current per unit cross-section area. We may again relate the direction of B to the sense of rotation of the current by use of the sign convention in which counterclockwise current is positive and B is positive when directed outward from the current loop.

6.4 *Some Properties of the Magnetic Induction Field*

In this section we summarize what we have learned about the magnetic induction field and add further to our understanding of its nature. As in the case of the electric field, we can represent magnetic field strength and direction by the convention of lines of force. This we connect with the inverse-square relationship that holds for current elements. We can again describe the magnitude of the magnetic field in terms of *density of lines* and may use this point of view for a quantitative measure of field strength. In the mks system a line of induction is called a *weber*. The unit of field strength is the weber per square meter. Another unit in common use is the *gauss*, where one weber per square meter is equal to 10,000 gauss.

In principle we are defining the units of **B** in terms of Eq. (6.2). There are, however, two difficulties with using this equation for this purpose. One is that it is a differential expression involving a length dl. The other is that we must choose units for the current i before evaluating the expression. We shall circumvent these diffi-

culties in the following way: As will be shown, we may define the unit of current, the *ampere*, in terms of the magnetic force between two conveniently chosen current loops and then use a convenient case, where Eq. (6.2) has been integrated, for determining B in terms of the current.

We first define unit current. We have already seen [by integrating Eq. (6.2)] that the field at a distance a from a long straight wire carrying current i is

$$B = \frac{\mu_0 i}{2\pi a}$$

where the lines of B form closed loops around the wire. If we take two parallel wires of length L, separated by a distance a, and form circular loops of each one, keeping the distance a small compared with the loop radius (see Fig. 6.11), this same equation holds for

Fig. 6.11 ***Force between two parallel current loops separated by a distance a.***

the field at one wire due to the current in the other. We have thus avoided the need for a very long pair of wires. We next apply Eq (6.3) for the force between current elements. This gives $dF = i'\,dl\,B$, where dl is an element of length of the wire loop on which we shall calculate the force. Substitution of $B = \mu_0 i/2\pi a$ and integration yields

$$F = \frac{\mu_0 i' i \int dl}{2\pi a} = \frac{\mu_0 i' i L}{2\pi a} \quad \text{newtons} \tag{6.13}$$

where L is the circumference of each loop. If we place the two coils in series, $i = i'$ and we can use the equation to allow for an experimental determination of unit current. Thus if we measure F in newtons and measure lengths in meters, i will be given in amperes. Once we are able to measure the current in amperes, as we could do with this pair of loops, any one of our calculations of the value of B in terms of i can be used for determining B. For example, again using a loop, we have seen that the field at its center is $B = \mu_0 i/2r$, where r is the radius of the loop.

We have thus been able to define magnetic field strength, using the force between two current circuits. However, the lines of magnetic field are *not* in the direction of the force on a current element. The vector cross product between $d\mathbf{l}$ and $\mathbf{B}$ gives the direction.

The experimental arrangement described above not only relates $\mathbf{B}$ to the current but also fixes the value of the unit for charge q in the mks system. This is done through the relationship given in Eq. (7 1), one ampere equals one coulomb per second. Through this relationship, the value of $\mu_0/4\pi$ is obtained, and the units in electrostatics are related to magnetic units.

We now turn to another contrasting property of the lines of magnetic induction. These lines are continuous and do not arise from any source in the way that lines of electric force originate on charges. Lines of magnetic induction thus form loops without any beginning or end. This property may be used to gain further insight into the nature of the magnetic field and also to allow for an extremely simple way of calculating the magnetic field in certain situations of high symmetry. Going back to the expression for the field around a long wire carrying a current i, we found that at a distance r from the wire, the field is given by

$$B = \frac{\mu_0 i}{2\pi r} \qquad \text{webers/m}^2$$

and that the magnetic lines are concentric around the conductor. Let us follow a particular line at a distance r around the wires and evaluate the line integral $\oint \mathbf{B}\cdot d\mathbf{l}$ around this path, starting at any point and returning to the same point. Since the path is everywhere parallel to $\mathbf{B}$ and since $B = \mu_0 i/2\pi r$ the integration becomes

$$\oint \mathbf{B}\cdot d\mathbf{l} = \frac{\mu_0 i}{2\pi r}\oint dl = \frac{\mu_0 i}{2\pi r}\,2\pi r = \mu_0 i \tag{6.14}$$

where i is the total current threading the path we choose. This is called *Ampère's circuital law.* While the integration was here performed over a circular path, Eq. (6.14) is true, whatever the nature of the closed path. This becomes apparent if any given path around the conductor is approximated by segments that either are

radial in direction or are circular arcs about the conductor (see Fig. 6.12). $\int \mathbf{B} \cdot d\mathbf{l}$ for a given radial segment depends only on the angle subtended at the wire by the arc, since for a given arc $\mathbf{B}$ falls

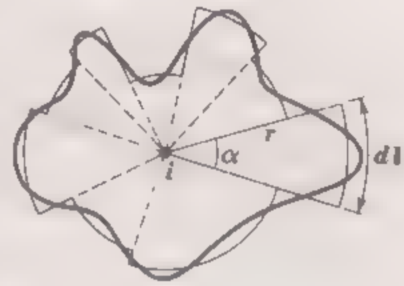


***Fig. 6.12** Approximation of an arbitrary closed path around a conductor by circular arcs and radial segments. Used for proof that $\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 i$ for any closed path enclosing a current i.*

off as $1/r$ while $d\mathbf{l}$ increases as r. The contribution of the radial segments to $\int \mathbf{B} \cdot d\mathbf{l}$ is zero, since $\mathbf{B}$ is everywhere $\perp$ to the radius vector. Therefore $\oint \mathbf{B} \cdot d\mathbf{l}$ is simply the value over all the circular segments, which is the same as over a circular path. When the current is not limited to a single wire, it can be made up of a distribution of many small current elements, each contributing its field $d\mathbf{B}$. The line integral can then be made up of the sum of elements $\oint d\mathbf{B} \cdot d\mathbf{l} = \mu_0 \, di$, which when added up give the same result as Eq. (6.14). Let us recall the somewhat similar result for the line integral of the electric field, $\oint \mathbf{E} \cdot d\mathbf{l} = 0$, which was obtained on the basis of conservation of energy. In the magnetic-field case, we have shown that the analogous result does not hold. However, when a path through which no current flows is chosen, as, for example, the path shown in Fig. 6.13, the line integral

$$\oint \mathbf{B} \cdot d\mathbf{l} = 0 \tag{6.15}$$

just as in the electrostatic case. Under these conditions (only) we

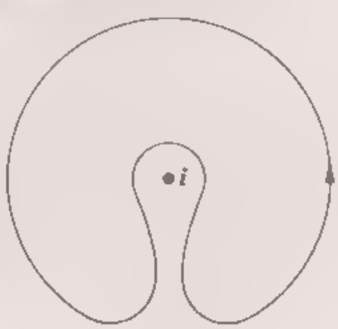


***Fig. 6.13** Path threaded by no current, for which $\oint \mathbf{B} \cdot d\mathbf{l} = 0$.*

can define a *magnetic potential* just as in the electric case. We shall not pursue this further, however.

6.5 Flux of Magnetic Induction

We introduce at this point a quantity of considerable importance, the *magnetic induction flux* Φ. This is the total number of lines of magnetic induction through a given area. The general equation for this is

$$\Phi = \int \mathbf{B} \cdot d\mathbf{S} \tag{6.16}$$

Since **B** gives the number of lines of **B** per unit area, where the area is taken perpendicular to the direction of **B**, this equation follows immediately from the definition of Φ. The quantity $d\mathbf{S}$ is an element of area (using vector notation) and is positive when directed *outward* (toward the viewer) from the area. The scalar or dot product takes account of the possibility that the direction of **B** may not be perpendicular to the area involved. In the special case that **B** is uniform and perpendicular to the area being considered, Eq. (6.16) reduces to $\Phi = BS$, where S is the area. We had a similar expression for the flux of electric field through a given area, $\int \mathbf{E} \cdot d\mathbf{S}$, though in the electric case there was no need to define a special symbol.

A consequence of the fact that lines of **B** close onto themselves and thus do not have sources, in contrast to the case of lines of **E**, is that

$$\int_{CS} \mathbf{B} \cdot d\mathbf{S} = 0 \tag{6.17}$$

for *any* closed surface.

6.6 Further Field Calculations

We next apply the line integral as developed in Sec. 6.4 to the calculation of the field due to two highly symmetrical current configurations, both of which we have calculated earlier by integrating the fundamental equation

$$d\mathbf{B} = \frac{\mu_0}{4\pi} \frac{i\, d\mathbf{l} \times \hat{\mathbf{r}}}{r^2} \tag{6.2}$$

a The magnetic field of a long straight conductor carrying a current i For this case we need only reverse the direction of the argument given in Sec. 6.4 We consider it here only to make the argument clear. We choose a concentric path around the conductor with a radius r. Knowing that $\mathbf{B}$ will be constant and parallel to the path around the entire loop, we can take $\mathbf{B}$ outside the integral and find

$$\oint \mathbf{B}\cdot d\mathbf{l} = B\oint dl = B2\pi r = \mu_0 i$$

or

$$B = \frac{\mu_0 i}{2\pi r}$$

our previous result.

b The field inside a long solenoid Here we start with a solenoid in the form of a torus as shown in Fig. 6.14 Let the average radius

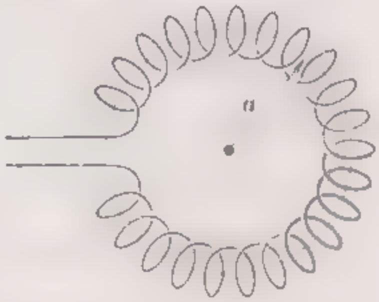


Fig. 6.14 Toroidal solenoid.

of the torus be a. Let there be N turns of wire carrying a current i. The length of the solenoid is $2\pi a$. We choose a path shown by the light line, over which we can see by symmetry the field $\mathbf{B}$ will be uniform and also parallel to the path. The total current threading this path is Ni, since each turn carries current in the same direction through the path we have chosen. The return current on the outside of the torus is outside the closed path and therefore need not be considered. The line integral becomes

$$\oint \mathbf{B}\cdot d\mathbf{l} = B\oint dl = B2\pi a = \mu_0 Ni$$

or $B = \dfrac{\mu_0 Ni}{2\pi a} = \dfrac{\mu_0 Ni}{L} = \mu_0 j^s$, the result we found earlier for a long straight solenoid, where j^s is the solenoidal current density. If we open out the toroid to form a straight solenoid, the field at its

center will not be seriously affected, though as we have seen, the field at the ends will be reduced by a factor of 2. Although our new method is only approximate, it does give us one piece of information that the detailed calculation did not give: The field at any point well inside the solenoid is the same, whether or not it is on the axis of the coil. In other words, the field inside a long solenoid is uniform far away from its ends.

6.7 *Torque on a Current Loop in a Uniform Magnetic Field. The Magnetic Dipole*

Now that we have finished our discussion of the production of magnetic fields by currents, we turn again to the study of forces on conductors carrying currents. We center our attention on a very simple shape—a rectangular loop as shown in Fig. 6.15. The axis

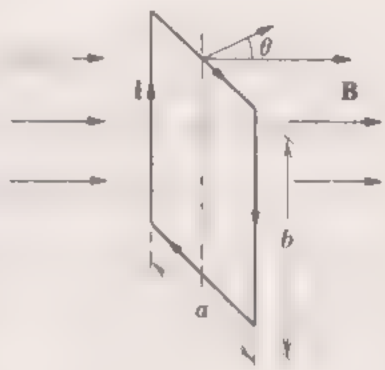


Fig. 6.15 Rectangular loop in a uniform magnetic induction field.

of the loop is perpendicular to the uniform magnetic induction **B**, and the normal to the loop makes an angle θ with the field direction. We shall find the torque on the loop and show that the net force on it is zero, in this special case of uniform field. A current i flows through the loop. If the current i is clockwise as viewed from the left in Fig. 6.15, then both i and **B** are negative (in the sense of our earlier definition of the sign convention for **B** and i) when we view the loop from the left and positive when we view from the right.

Using the force equation (6.3), applied to the top and bottom sides, we find

$$\mathbf{F} = (i\mathbf{a} \times \mathbf{B}) \qquad \text{or} \qquad F = iaB \sin\left(\frac{\pi}{2} - \theta\right) = iaB \cos\theta \tag{6.18}$$

When the vector cross-product rule is applied, this force is upward on the top side and downward on the bottom side. The net force on

the rigid loop from these elements is thus zero, as is the torque. Applying the same equation to the vertical sides, we find

$$F = ibB \tag{6.19}$$

where the forces are again equal and opposite and outward but have components perpendicular to the plane of the coil. The torque

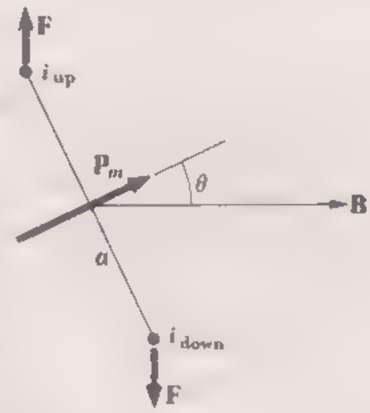


***Fig. 6.16** Top view of a rectangular current loop in a uniform magnetic induction field. The normal to the plane of the loop makes an angle θ with the magnetic field.*

on the coil is thus not zero. Using Fig. 6.16, which is a view looking downward on the loop, we find for the total torque

$$\tau = ibB \sin\theta = iBA \sin\theta \tag{6.20}$$

where A is the area ab of the loop. If there are N turns on the loop, the equation becomes

$$\tau = NiAB \sin\theta \tag{6.21}$$

If we let

$$NiA = p_m \tag{6.22}$$

as in Sec. 6.3, where p_m represents the magnetic dipole moment, the equation becomes

$$\tau = p_m B \sin\theta \tag{6.23}$$

The magnetic dipole moment p_m may be treated as a vector quantity pointing in the direction of the magnetic induction flux at the center of the current loop. From this point of view, Eq. (6.22) gives the direction as well as the magnitude of p_m. Thus if we view the current loop NiA from the direction for which the current flows counterclockwise (positive), the magnetic induction B due to the current and hence also the dipole moment are directed outward (toward the viewer). The detailed analysis we have given

shows that the torque due to an externally applied magnetic induction field tends to align the magnetic dipole moment parallel to the external field. This behavior is the same as in the electric case, in which the electric dipole, treated as a vector quantity pointing from − to + ends of the dipole, tends to line up parallel to an external electric field.

***Fig. 6.17** Current loop made up of many elementary rectangular loops. Effects of all currents except those around the periphery cancel.*

The justification for using p_m for representing NiA is that whatever the shape of the loop, the torque depends only on NiA and not on the shape. We can see this by taking any arbitrary shape of loop and approximating it as a sum of rectangular loops as shown in Fig. 6.17. If the same current is sent through each rectangular loop in the same sense, the inside currents will just cancel each other, leaving only the peripheral current, which approximates the original loop as closely as we wish. The total torque is then the sum of the separate torques on individual rectangles and so must depend only on the current and total area of the loop.

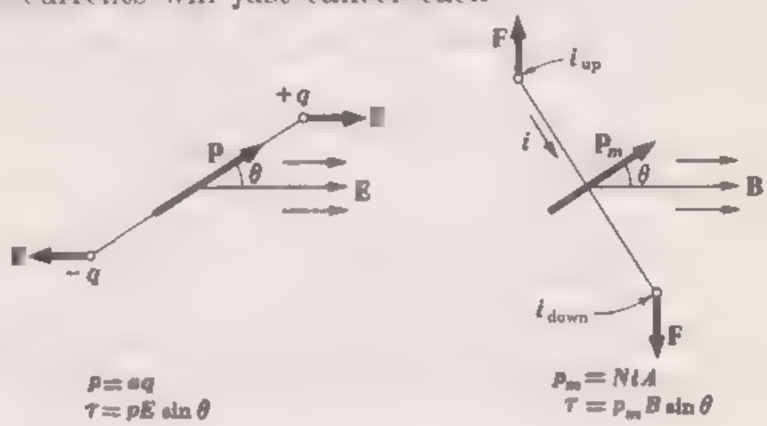


***Fig. 6.18** Electric and magnetic dipoles in uniform electric and magnetic fields.*

Figure 6.18 calls attention to the close parallel between an electric dipole in an electric field and a current loop in a magnetic field. In order to make the two situations similar, we must choose for the axis of the dipole moment of the loop the direction of the

normal to the plane of the loop. The magnetic dipole moment points in the direction from which the loop current is counterclockwise.

Actually, the similarity in form between these very different quantities, the electrostatic and the magnetic dipole, goes even further. We have found the magnetic field along the axis of a magnetic loop or dipole, which we can write as

$$B = \frac{\mu_0}{4\pi}\frac{2p_m}{r^3} \tag{6.9}$$

Earlier we found for the electric field along the axis of an electric dipole

$$E = \frac{1}{4\pi\epsilon_0}\frac{2p}{r^3} \tag{2.5}$$

Thus the expressions are very similar. Not only is the comparison valid along the axis of the dipole but the same similarity holds at all points in space around the dipole as long as we stay far enough away so that the mathematical approximations used for the electrostatic case are valid. It is clear that at distances comparable to the dimensions of the dipoles the fields are of different shapes. For comparison we give below the similar electric and magnetic expressions that apply far away, in terms of the tangential and radial components of the fields. We have not proved the equations for the magnetic case but shall quote them anyway.

$$\begin{aligned} E_r &= \frac{1}{4\pi\epsilon_0}\frac{2p}{r^3}\cos\theta \qquad & B_r &= \frac{\mu_0}{4\pi}\frac{2p_m}{r^3}\cos\theta \\ E_\theta &= \frac{1}{4\pi\epsilon_0}\frac{p}{r^3}\sin\theta \qquad & B_\theta &= \frac{\mu_0}{4\pi}\frac{p_m}{r^3}\sin\theta \end{aligned} \tag{6.24}$$

The angle θ in these equations is the angle between the dipole moment and the field direction (see Fig. 6.18).

6.8 Forces on Isolated Moving Charges

Our next task, to inquire about magnetic forces on individual moving charges, is comparatively easy, since we have already examined forces on current elements and these are made up of a flow of individual charges. Our problem is to modify

$$d\mathbf{F} = i\,d\mathbf{l} \times \mathbf{B} \tag{6.3}$$

to allow it to apply to an individual charge e moving with a velocity

v. In order to do this, we look ahead to one of the results in the next chapter that gives current in terms of the motion of a group of charges. Equation (7.5) gives the expression for the current i in terms of N charges of charge e per unit volume moving with an average drift velocity $\mathbf{v}$ through a conductor of cross-section area A.

$$i = NevA \tag{7.5}$$

Substituting this in the equation for the force on a current element [Eq. (6.3)], we get

$$d\mathbf{F} = NeA\ dl\ \mathbf{v} \times \mathbf{B}$$

This is the force on a total charge $NeA\ dl$. Dividing by $NA\ dl$, we have the force on a single charge e:

$$\mathbf{F} = e(\mathbf{v} \times \mathbf{B}) \tag{6.25}$$

The force is in newtons when the charge is measured in coulombs, the velocity in meters per second, and $\mathbf{B}$ in webers per square meter.

We next investigate some consequences of this force on individual charges. We could have started with this last equation as the fundamental force-field relationship and then worked backward to the force equation for a current element. Our choice was essentially arbitrary.

The first and most obvious conclusion from Eq. (6.25) is that a magnetic field acting on a moving charged particle cannot change the speed or kinetic energy of the particle. This follows from the nature of the vector cross product, which makes the resultant force be always perpendicular to the direction of motion. This means that the magnetic field can do no work on a charged particle, although it can change its direction of motion.

A further result of the mutually perpendicular magnetic force and velocity vectors is that in a uniform magnetic field the motion of charged particles is circular or helical. Consider first a particle of positive charge q moving in a plane perpendicular to a uniform magnetic field, which in Fig. 6.19 is shown pointing into the paper.

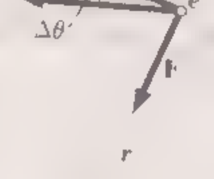


***Fig. 6.19** Path of a moving charged particle in a perpendicular magnetic field.*

In a time Δt it acquires a transverse velocity Δv, which can be obtained from $\mathbf{F} = q(\mathbf{v} \times \mathbf{B})$, using the equation $F\,dt = m\,dv$:

$$\Delta v = \frac{qvB}{m}\,\Delta t$$

The angular change in the direction of v is given by

$$\Delta\theta = \frac{\Delta v}{v} = \frac{qB}{m}\,\Delta t$$

or the angular velocity

$$\omega = \frac{\Delta\theta}{\Delta t} = \frac{qB}{m} \tag{6.26}$$

Since $\mathbf{v}$ remains constant in magnitude and the velocity uniformly changes direction, the orbit must be a circle. The radius of curvature can be calculated from Newton's law, $F = ma = mv^2/r$, where the centripetal force is supplied by the magnetic field. Thus we have

$$\frac{mv^2}{r} = qvB \qquad \text{or} \qquad r = \frac{mv}{qB} \tag{6.27}$$

In the more general case where the velocity is not limited to the plane perpendicular to $\mathbf{B}$, Eq. (6.25) shows that the component of velocity in the perpendicular plane will behave as shown in Eq. (6.27) while the component parallel to $\mathbf{B}$ will be unaffected. If there is an additional velocity component parallel to $\mathbf{B}$, the orbit is helical.

A final important consequence of the nature of the magnetic force on moving charges is that in a uniform field the frequency of rotation of a moving charge is constant, independent of the velocity of the charge. We have already seen this in the result of Eq. (6.26). It is also apparent from Eq. (6.27) if we substitute $r\omega = v$ and solve for ω, giving again the result of Eq. (6.26). This is often called the *cyclotron equation* for reasons discussed later.

We now discuss briefly the difficulty mentioned earlier (Sec. 6.2), the net force due to the magnetic interaction between two moving charges. When two charges are moving along parallel lines, the forces of interaction can be shown to be equal and opposite, as we saw in the case of forces between current elements. If, however, the motion is not along parallel paths, there will appear

to be a net force on the system, as suggested by our earlier discussion and shown in Fig. 6.7. This apparent breakdown of Newton's third law is resolved when we take account of an additional momentum that must be associated with the *electromagnetic* field of the moving charges. When this additional momentum is considered, the total momentum of a system of two or more moving charged particles in the absence of external fields is found to be constant.

6.9 Applications

In this section we discuss a number of important applications of the magnetic-force law on moving charged particles.

A velocity selector In a large class of experiments in which the motion of charged ions or electrons is to be studied, it is important to have a source of particles, all having the same velocity. Since most sources of electrons or ions emit particles with a wide range of velocities, a velocity selector is often essential. It is easy to understand the behavior of such a selector (Fig. 6.20), which uses both

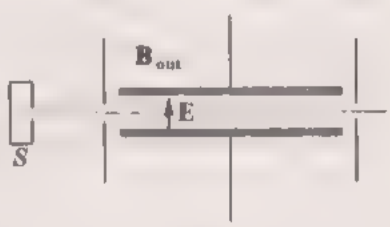


Fig. 6.20 *Velocity selector involving crossed* **E** *and* **B** *fields.*

electric and magnetic forces. A capacitor-like arrangement provides a uniform electric field **E**, upward in the plane of the paper. A uniform **B** field is provided perpendicular to the paper. Collimated charged particles emerging from the slit in the source S reach the output slit only if the electric and magnetic forces acting are equal and opposite. This requires

$$Eq = qvB \tag{6.28}$$

for the simple arrangement involved. Thus only charged particles with a velocity

$$v = \frac{E}{B} \tag{6.29}$$

can come out of the slit.

Determination of e/m In building up our knowledge of atomic structure, the determination of e/m for electrons and for ions has been a basic tool. Such determination is made possible by taking ions after passage through a velocity selector through a circular path perpendicular to a uniform magnetic field. Using an appropriate detector of ions and a slit system, we can determine the radius of curvature of the electron or ion path. Knowledge of the velocity then allows evaluation of e/m. Thus, since $r = mv/eB$ [Eq. (6.27)] and $v = E/B$ [Eq. (6.29)], we find

$$\frac{e}{m} = \frac{E}{rB^2} \tag{6.30}$$

J. J. Thomson used this method for the measurement of e/m for electrons in 1897 at the Cavendish Laboratory of Cambridge University in England. As used for the study of charged ions, such an apparatus is called a *mass spectrograph*. Studies of e/m for charged ions have elucidated many of the important facts of atomic structure, including the existence of isotopes.

The cyclotron Probably the most familiar of all the machines for accelerating charged particles (ions) to high velocity, the cyclotron takes advantage of electric and magnetic forces on the ions. Ions starting from a central source are caused to move in circular paths by a magnetic field perpendicular to their motion. Motion takes place within two hollow electrodes between which an alternating voltage is applied. Ions that start out at the right time feel an accelerating electric field each time they pass from one electrode to the other. Such ions make larger and larger orbits as they gain kinetic energy. However, they continue to stay in step with the alternating voltage, since their angular frequency is constant [Eq. (6.26)]. A sketch of a cyclotron is shown in Fig. 6.21.

For energies greater than about 20 million electron volts (Mev) the ordinary cyclotron fails because of the relativistic increase in the mass of the particles being accelerated. Thus the frequency of rotation of charged particles as given by Eq. (6.26) changes at high energies, and the particles do not remain in synchronism with the alternating voltage applied. Machines have been developed that avoid this difficulty by appropriately modifying the magnetic field or the frequency of the accelerating voltage as the particles increase in energy. Machines such as the proton synchrotron produce

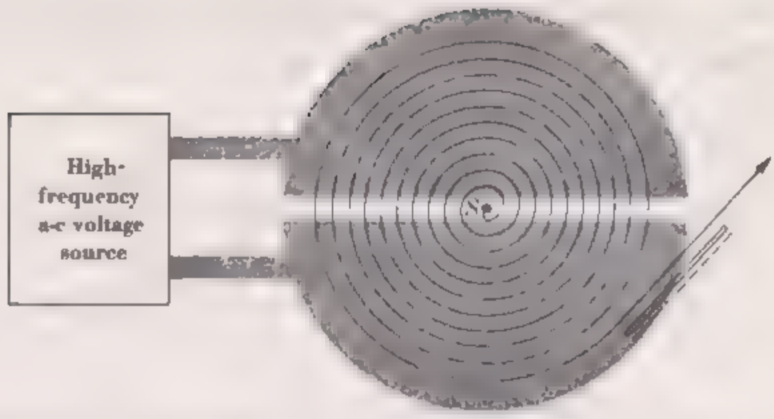


Fig. 6.21 Cyclotron. Top view of cyclotron electrodes placed in an evacuated chamber between the poles of an electromagnet. Positive ions emitted by the source S travel in circular orbits inside the hollow electrodes perpendicular to the magnetic field. Each time the ions traverse the gap between the electrodes they are accelerated by a potential difference due to an applied alternating voltage synchronized with the ion motion. As the ions gain energy, the radius of their path increases, until they are brought out of the magnetic field region by a negatively charged deflector plate.

particles with energies in the 1 to 10 billion electron volt (Bev) range.

Hall effect The Hall effect relates to the generation of a voltage when a current-carrying conductor is placed in a magnetic field. Useful information regarding current carriers in conductors is obtainable from this effect. Figure 6.22 shows the arrangement

Fig. 6.22 Geometry involved in the determination of the transverse Hall effect.

used in the transverse Hall effect. If placed in a field B_y, carriers of a current i_x are deflected (as shown here for electrons) until a transverse field E_z is built up by the deflected charge. This is known as the Hall field and is obtained from the equation

$$eE_z = e|\mathbf{v} \times \mathbf{B}| = evB \tag{6.31}$$

If we write j, the current density (current per unit cross-section

area as defined in Sec. 7.2) in terms of the density of carriers N, their charge e, and their mean velocity v,

$$\mathbf{j} = Ne\mathbf{v} \tag{7.4}$$

we may express the Hall constant R_H as

$$R_H = \frac{E_z}{jB} = \frac{1}{Ne} \tag{6.32}$$

This constant may thus be determined experimentally by the measurement of the *Hall voltage* (from which the value of E_z can be calculated), current density, and magnetic field. The sign and magnitude of the Hall constant give the sign of the carriers of current and their density. For most metals the carriers are found to be electrons, and for many metals the density of carriers is in good agreement with the number of valence electrons in the atoms making up the metal. However, there are cases where the carriers are positive (e.g., Be, Zn, Cd), and for Bi a very large value of the Hall coefficient suggests an anomalously low concentration of electrons (~0.004 electron per atom). These surprising results are well understood qualitatively and are associated with the quantum nature of solids. Another important result from Hall-effect measurements on metals is that N, the charge carrier density, varies only slightly with temperature.

PROBLEMS

6.1 Two long parallel wires are separated by 10 cm, and each carries 10 amp in the same direction. Calculate the force between the wires per unit length.

6.2 Find the magnetic induction field at the center of a square wire loop of side length 10 cm, carrying 10 amp.

6.3 A conductor of circular cross section of radius a carries a current of uniform current density j. Find the magnetic induction field at all distances r from the center of the conductor.

6.4 In a conductor such as in Prob. 6.3, is there a magnetic force that tends to concentrate the current and make j nonuniform? Describe. In a conductor, is there any other force that would tend to stabilize the current density and prevent a concentration of the current?

6.5 A long wire has a semicircular loop of radius r as shown in Fig. P6.5. A current i is flowing. Find the magnetic induction field at the center of curvature of the loop.

Fig. P6.5

6.6 A long, thin conductor of width b carries a current of i amp. Find the magnetic induction field in the plane of and outside the conductor at a distance a from its near edge.

6.7 The coaxial line shown in Fig. P6.7 carries the same current i up the inside conductor of radius a as down the outer conductor of inner radius b and outer radius c. Find the magnetic induction field at all distances r from the center of the conductor. (Use Ampère's circuital law.)

Fig. P6.7

6.8 A wire 10 cm long can slide on two parallel rods tipped at 45° to the vertical as shown in Fig. P6.8. These rods connect to a source of emf that produces a current i_1 between the two ends of the rod. Another long conductor parallel to the movable wire carries a current i_2. The currents i_1 and i_2 are in opposite directions. If the weight of the movable rod is mg, find the equilibrium distance between the two current-carrying wires.

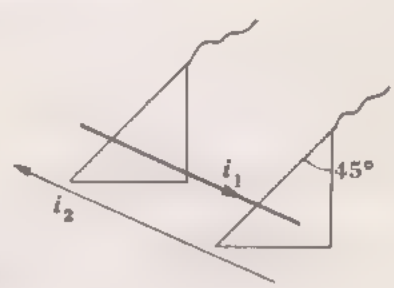


Fig. P6.8

6.9 A solenoid 20 cm long of radius 2 cm is wound uniformly with 3,000 turns of wire. A current of 2 amp flows through the coil.

a What is the solenoidal current density j^s?

b What is the value of B on the axis of the solenoid, at the middle?

c What is the value of B on the axis at an end?

d What is the flux Φ through the coil at the middle?

e What is the flux Φ through one end?

6.10 An insulating circular disk of radius a has a uniformly distributed static charge of σ coulombs/m^2. The disk rotates about its center with an angular velocity ω. Find the magnetic induction field at its center.

6.11 Find the magnetic induction field at a distance b from the rotating disk of Prob. 6.10, along its axis of rotation, for $b \gg a$.

6.12 Two circular coils of N closely wound turns of radius a are coaxial and are separated by a distance b. Find the force between the two coils when a current i passes through each coil.

6.13 The axis of a circular coil of radius 10 cm makes an angle θ with a uniform field B. The coil has 10 turns and carries a current of 5 amp. Find the torque on the coil.

6.14 A toroidal coil of inner radius 5 cm and outer radius 6 cm is uniformly wound with 1,000 turns and carries a current of 5 amp.

a Find the average magnetic induction field in the coil by using Ampère's circuital law.

b Find the magnetic induction field in the coil at a distance of 5.1 cm from the axis of the toroid.

6.15 A current of 10 amp flows in the long wire shown in Fig. P6.15. Find the total flux through the area *abcd*.

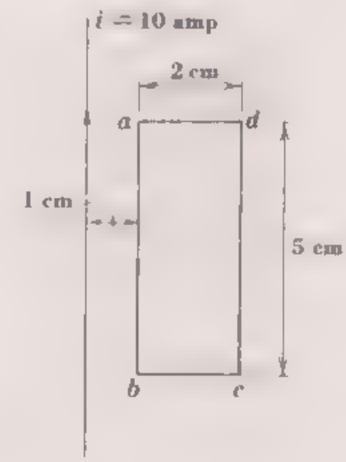


Fig. P6.15

6.16 In Fig. P6.15 a wire loop carries 5 amp around the path *abcd* in a clockwise direction. Find the magnitude and direction of the force acting on the loop due to the magnetic field of the upward current of 10 amp in the straight wire.

6.17 In a transverse-Hall-effect experiment, a current of 10 amp flows through a conductor of square cross section 0.5 cm on a side. The Hall voltage induced by a magnetic field of 2 webers m^2 is 2.5×10^{-6} volt. Calculate the Hall constant R_H. If the carriers of current

are electrons, find N, the density of carriers in the conductor, and draw a sketch showing the relative directions of the current, the magnetic induction field B, and the Hall field E_z.

6.18 What would be the transverse Hall field for a conductor having an equal density of positive and negative mobile charge carriers?

6.19 The steady magnetic field of a cyclotron has a value of 0.5 weber/m^2. What must be the frequency of the voltage variation on the two electrodes in order that the alternation of their potentials be synchronous with the motion of hydrogen ions? What will be the energy in joules and in electron volts of the ions when the radius of their path is one meter? How many revolutions will be required for the ions to gain this energy if the maximum potential difference between the electrodes is 20,000 volts?

6.20 A hydrogen atom consists of a proton and an electron separated by a distance of 0.5×10^{-10} m. Assuming that the electron moves in a circular orbit around the proton with a velocity of 10^{13} cps, find the magnetic induction field at the nucleus due to the moving electron.

6.21 The Helmholtz arrangement of two coils provides a large region of uniform field. Two similar coils carrying the same current are placed on the same axis, as shown in Fig. P6.21, separated by a distance equal to the coil radius. Show that at point P on the axis, this arrangement gives both dB/dx and d^2B/dx^2 equal to zero.

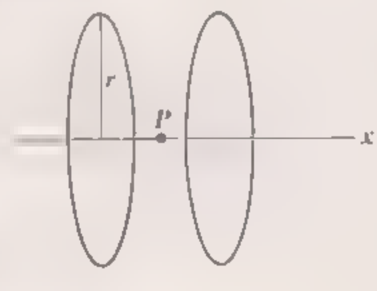


Fig. P6.21

6.22 Taking the radius of the earth as 6,400 m and the horizontal component of the earth's magnetic field at the equator as 0.4×10^{-4} weber/m^2, with what minimum velocity could a proton encircle the earth at the equator?

6.23 An ion in vacuum starts from rest and is accelerated between two parallel plates having a potential difference of 1,000 volts as shown in Fig. P6.23. On emerging from between the plates, the ion moves into a uniform magnetic induction field of 0.1 weber/m^2, directed perpendicular to the path of the ion. If the radius of curvature of the ion path is 0.3 m, what is the mass of the ion if it is singly charged (has one electronic charge)?

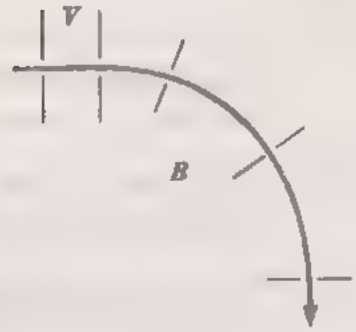


Fig. P6.23

6.24 Compute the tension in a circular loop of flexible wire of radius a, carrying a current I and lying in a uniform magnetic field of flux density B perpendicular to the plane of the loop.

6.25 Show that the net force on an element of current in a uniform magnetic induction field depends only on the positions of its two ends and not on the shape of the element.

6.26 A current-carrying circular loop of radius r lies in a divergent magnetic induction field whose lines of B make an angle θ with the plane of the loop, as shown in Fig. P6.26. If the loop has N turns carrying a current I_0 in a clockwise direction as seen from above and if the external B field in the plane of the loop is B_0, find the magnitude and direction of the force on the loop.

Fig. P6.26

6.27 Find the direction and magnitude of the magnetic field at the two points a and b (a is at the center of curvature of the semicircular loop of radius r, and b is midway between the two wires) as shown in Fig. P6.27.

Fig. P6.27

6.28 A large number N of closely spaced turns of fine wire are wound in a single layer upon the surface of a wooden sphere with the planes of the turns perpendicular to an axis of the sphere and completely covering its surface. The current in the winding is i. Find the magnetic flux density at the center of the sphere.

6.29 A pair of infinitely long thin wires carry equal currents i. They are bent as shown in Fig. P6.29. What is the magnetic induction B at the center of the circular parts?

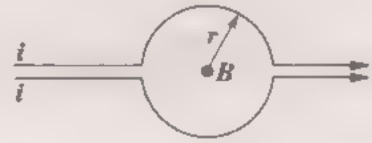


Fig. P6.29

6.30 Electrons are revolving in a uniform magnetic field. What is the value of the field B such that the electrons make one complete revolution in 10^{-8} sec?

6.31 A flat coil is wound so that it contains a very large uniform number of turns per unit distance along its radius, as shown in Fig. P6.31.

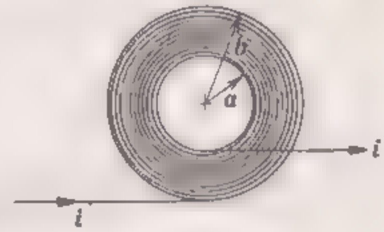


Fig. P6.31

If this number is Z, show that the magnetic induction at the center of such a coil of inside radius a, outside radius b, carrying a current i, is given by

$$B = \frac{\mu_0}{2} iZ \ln \frac{b}{a}$$

SEVEN

Current, Resistance, and Electric Circuits

7.1 Introduction

In this chapter we discuss electric circuits and the factors that control currents in circuits. After considering resistance and Ohm's law, we study the dissipation of energy when current flows in a resistance and discuss electromotive force as a source of electric energy. We develop circuit equations and study methods of measuring current and voltage. Finally, we depart from fundamental electrical theory to discuss the causes of resistivity in metals, the mechanism by which chemical cells produce electric energy, and the phenomena of contact potential and thermal electromotive force.

7.2 Electric Current

In this section we limit the discussion to the simplest ideas about currents and metallic conduction. We postpone a more detailed examination of the reasons for conduction in metals and the nature of the metallic state.

Basically, electric current is simply a flow of charge. If there is a surface through which charges are flowing steadily, such as a cross section cut through a wire, the current i is defined as

$$i = \frac{dq}{dt} \qquad \text{coulombs/sec} \tag{7.1}$$

A current of one coulomb per second is defined as one ampere in both the mks and the practical system of units (see Chap. 15). The positive direction of current is taken conventionally as the direction of flow of positive charge. In a majority of cases it is actually a negative charge flow in the opposite direction that produces current.

When the rate at which charge is passing by varies over the surface, we can more conveniently discuss the current density $\mathbf{j}$, which is related to the current by the equation

$$i = \int \mathbf{j} \cdot d\mathbf{S} \qquad \text{amp} \tag{7.2}$$

where $d\mathbf{S}$ is an element of the area A. The integral is taken over the surface through which we are calculating the current, and the unit of current density is the ampere per square meter. Use of vector notation allows the consideration of a cross-section area not perpendicular to the current flow. Thus in Fig. 7.1 the integral

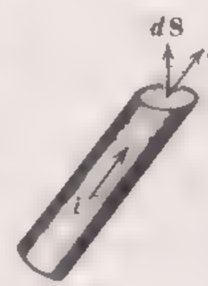


***Fig. 7.1** Relationship between current and current density, $i = \int \mathbf{j} \cdot d\mathbf{S}$.*

allows for the angle between $d\mathbf{S}$ and $\mathbf{j}$. The value of the integral is independent of the shape of surface taken and thus gives the same current through the wire as long as the surface taken cuts completely across the region of charge flow. When the current density is uniform, Eq. (7.2) can be integrated to give

$$i = \mathbf{j} \cdot \mathbf{A} \tag{7.3}$$

When the area is taken perpendicular to the current density, this becomes $i = jA$, where A is the cross-sectional area of the conductor.

Current is due to the drift of charge carriers with a mean drift velocity $\mathbf{v}$, each carrying a charge e and having a density of N carriers per unit volume, so

$$\mathbf{j} = Ne\mathbf{v} \tag{7.4}$$

or, if A is the cross-sectional area of the conductor,

$$i = NevA \tag{7.5}$$

This is shown using Fig. 7.2. We ask for the rate at which charge crosses the area A. In a time dt a cylinder of charge of length $\mathbf{v}\,dt$

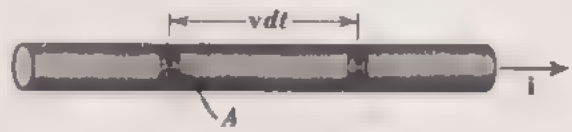


Fig. 7.2 Calculation of current in terms of drift velocity.

and area A passes through the surface we have chosen. This volume is $A\mathbf{v}\,dt$ and it contains $NA\mathbf{v}\,dt$ charge carriers. Since each carrier has a charge e, the total charge flow in the time dt is $dq = Ne\mathbf{v}A\,dt$. Evaluation of $dq/dt = i$ gives the result of Eq. (7.5).

7.3 Resistance, Ohm's Law

Equation (7.5), which shows that the current is proportional to the mean drift velocity, provides the clue for understanding Ohm's law, the law of proportionality between current flow and potential difference along a metallic conductor. Although the individual conduction electrons in a metal move at quite high velocities ($\sim 10^6$ m/sec), this motion in the absence of an externally imposed electric field is completely random and results in no net flow of charge. We may think of this random motion as resulting from random elastic collisions between the electrons and the metal lattice (ion cores). The average time between collisions is known as the *mean free time*, and the average distance traveled between collisions is called the *mean free path*. This is the same terminology used in discussing the random motion of molecules in a gas. Because of the random nature of this motion, on the average no net charge is transferred. However, if an electric field is applied to the metal, the path of each electron is bent in its random motion in the direction of the force produced by the field. Figure 7.3 is a crude picture of the effect of

the external field in causing the electron paths to curve in the direction of the force to produce a net drift velocity. This curvature results from the acceleration of the charge between collisions. We may assume that on each collision with the lattice, the excess energy

Fig. 7.3 ***Effect of external field on random electron motion.***

picked up by the electron in the external field is lost to the lattice. We now can associate the mean drift velocity v with the acceleration due to the externally applied electric field. Figure 7.4 shows a

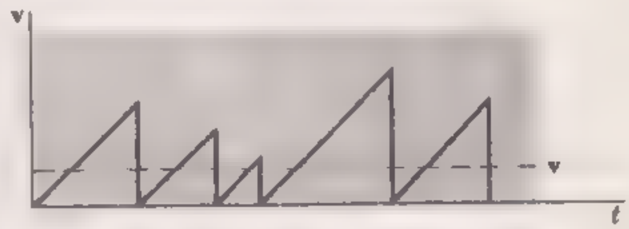


Fig. 7.4 ***Time variation of velocity produced by external field. The accumulated velocity goes to zero at each collision with lattice.***

qualitative picture of the way in which the excess velocity produced by the external field varies with time. During each free path the electric field produces a uniform acceleration in the field force direction ($\mathbf{F} = e\mathbf{E}$). Each collision reduces the accumulated velocity produced by the external field to zero, after which the acceleration again causes a uniform velocity increase until the next collision with the lattice. The drift velocity is the time average of this motion superposed on the generally much higher random velocity of the electrons.

Since the drift velocity is always much less than the random velocity of the electrons, the mean time between collisions is independent of the applied field. Thus the only factor influencing the drift velocity is the acceleration due to the electric field force $\mathbf{E}e$ on the electrons. Since the current is proportional to the drift velocity, we have at once

$$i \propto E \tag{7.6}$$

which is the result we needed in order to understand Ohm's law. The current is proportional to the force on the electrons imposed by the electric field because of the effect of the collisions between electrons and lattice. The situation is similar to viscous flow, such as, for example, the velocity of fall of raindrops in air under gravitational force. The raindrops fall with a constant velocity that depends upon the gravitational force acting on them. If, instead, the raindrops were in free fall, their velocity would increase continuously, as would the current in the absence of the viscous effect of collisions.

For the simple case of a conductor of uniform cross section and length L, an applied potential difference V between the ends gives rise to a uniform field given by $E = V/L$. For this conductor, we may replace the field E in the proportionality equation (7.6) by V/L to get

$$i \propto V$$

or

$$V = iR \tag{7.7}$$

where R, the resistance, gives the proportionality between V and i. This is the usual expression of Ohm's law. With V in volts and i in amperes, R is in ohms in both the mks and practical systems of units (see Chap. 15).

Although electric fields in metals vanish in equilibrium, the present case is not one of equilibrium, but of steady state. That is, charge carriers (electrons, in the case of metals) are continually moving under the influence of the field and are prevented from producing a zero field situation by the potential difference maintained between the two ends of the conductor.

We may relate the resistance R to the properties of the conduction electrons in the metal. Using the same simple shape of a conductor of length L and cross-section area A, we calculate the value of R using the ideas of mean free time, etc., as discussed above. Using Eq. (7.5), $i = NevA$, we evaluate v, the mean drift velocity. If 2τ is the mean time between collisions,

$$v = \frac{1}{2} eE \frac{2\tau}{m} \tag{7.7a}$$

This follows from $F = eE = ma$, where $v = a(2\tau)$ is the accumulated velocity after a time 2τ. The factor ½ is due to the fact that the mean velocity during each free path is ½ the maximum velocity during that interval. τ is called the *relaxation time*. If we replace E by V/L, we find

$$i = \frac{Ne^2\tau A}{mL} V \tag{7.8}$$

This gives

$$R = \frac{m}{Ne^2\tau}\frac{L}{A} = \rho\frac{L}{A} \tag{7.9}$$

The quantity ρ as defined by Eq. (7.9) is called the *resistivity*. It is equal to the resistance of a conductor of unit cross section and unit length. Thus,

$$\rho = \frac{m}{Ne^2\tau} \qquad \text{ohm-m} \tag{7.10}$$

The *conductivity* σ is the reciprocal of the resistivity and is given by

$$\sigma = \frac{Ne^2\tau}{m} \qquad (\text{ohm-m})^{-1} \tag{7.11}$$

The resistivity depends only on the behavior and number of conduction electrons and not on the shape of the conductor.

Substitution of known quantities in Eq. (7.10) shows that for metals at room temperature, τ is of the order of 10^{-10} seconds. Very pure metals at very low temperatures may have values of τ as long as 10^{-7} seconds. This result is discussed in Section 7.10.

Another form of Ohm's law is obtained by starting with Eq. (7.6), $i \propto E$, which states that current through a conductor is proportional to the electric field. We may express this same proportionality in terms of the current density $j(= i/A)$ and write

$$\mathbf{j} \propto \mathbf{E}$$

The proportionality constant between current density and electric field is σ, the conductivity, so we can write

$$\mathbf{j} = \sigma\mathbf{E} \tag{7.12}$$

This is a vector equation since both $\mathbf{j}$ and $\mathbf{E}$ are vector quantities. We can show that this is exactly equivalent to our first expression for Ohm's law [Eq. (7.7)] by considering a length L of a conductor of uniform cross section A as shown in Fig. 7.5. Let the potential difference across the length L be V and write

$$i = \frac{V}{R} = \frac{\sigma V A}{L}$$

where we have replaced R by $\frac{1}{\sigma}\frac{L}{A}$, as required by Eqs. (7 9) and (7.11). When we solve this for the current density $\mathbf{j}$ and replace V/L by $\mathbf{E}$, we get the alternative form given by Eq. (7.12).

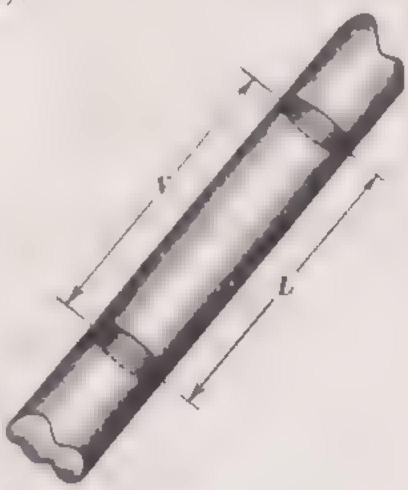


***Fig. 7.5** Length of conductor L with cross section A and applied voltage V. Current density is given by* $\mathbf{j} = \sigma\mathbf{E}$*, or current is given by* $i = E/R$*, where* $R = \frac{1}{\sigma}\frac{L}{A}$

As long as the conducting material is isotropic, the directions of $\mathbf{E}$ and $\mathbf{j}$ are the same. The advantage of this form of Ohm's law is that it is a *microscopic* expression. That is, it tells the relationship between current density and field at each point in a conducting medium. This contrasts with $V = iR$, which gives the total current through a finite body of resistance R across which a potential difference V is maintained.

When we have a number of resistors of resistance R_1, R_2, R_3, . . . connected in series, they can be shown to be equivalent to a single resistor of resistance $R = R_1 + R_2 + R_3 + \cdots$. Thus, in

series connection the current through each resistor is the same, so we have for the potentials across each:

$$\begin{aligned} V_1 &= iR_1 \\ V_2 &= iR_2 \\ V_3 &= iR_3 \\ &\cdots \end{aligned}$$

Adding these to get the total voltage V across the series array, we find

$$V = i(R_1 + R_2 + R_3 + \cdots) = iR$$

so the equivalent resistance is found to be the sum of the individual resistances.

In the case of parallel combinations, the voltage V is the same across each resistor. Thus $V = i_1R_1 = i_2R_2 = i_3R_3 \ldots$.

$$\begin{aligned} i &= i_1 + i_2 + i_3 + \cdots \\ &= \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3} + \cdots = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots\right) = V\frac{1}{R} \end{aligned}$$

Thus the reciprocal of the equivalent resistance is equal to the sum of the reciprocals of the resistances.

These results may be written as follows:

Resistances in series:

$$R = \sum_i R_i \tag{7.13}$$

Resistances in parallel:

$$\frac{1}{R} = \sum_i \frac{1}{R_i} \tag{7.14}$$

When the cross section of a conducting body is nonuniform, the current density varies through the body and the problem of calculating its resistance becomes more complicated than in the case of a simple body of uniform cross section, where we can use $R = \rho L/A$ [Eq. (7.9)]. For all but the simplest shapes, the mathematical difficulties can become rather serious, but we shall take a very simple example to show the method of attack. The problem is mathematically identical with the calculation of heat flow through a thermally conducting body.

Consider the radial flow of current through a circular slab of conducting material from some radius a at a potential V_a to the outer radius V_b (Fig. 7.6). Let the conductivity of the circular

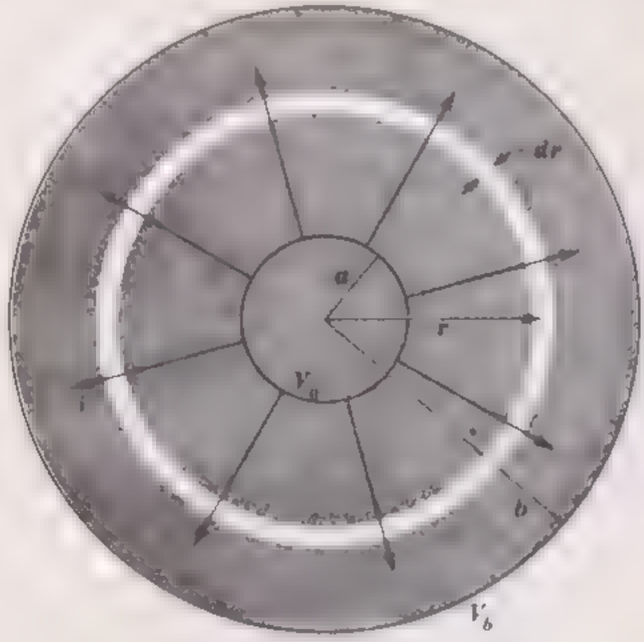


***Fig. 7.6** Calculation of resistance for radial current flow through a circular disk. Flow is from radius a at V_a to radius b at V_b.*

slab be σ and its thickness be t. Since the cross section of this conductor perpendicular to the current flow is $2\pi rt$ at any radius r and therefore varies as we move out from one end of the conductor to the other, we do not expect the voltage gradient or electric field to be uniform. However, the total current through any cross section is the same as that through any other cross section, so we can write an expression for the current,

$$i = jA = \sigma AE = -\sigma 2\pi rt \frac{dV}{dr} \tag{7.15}$$

where j is the current density at any radius r. When we solve this for dV across a ring of width dr, we get

$$dV = -\frac{i}{2\pi t\sigma}\frac{dr}{r}$$

The total voltage difference between the two ends of the conductor is

$$V_b - V_a = \int_a^b dV = -\frac{i}{2\pi t\sigma}\int_a^b \frac{dr}{r} = \frac{i}{2\pi t\sigma}\ln\frac{r_a}{r_b}$$

The resistance of the conductor is then given by

$$R = \frac{V_b - V_a}{i} = \frac{1}{2\pi t\sigma} \ln \frac{r_a}{r_b} \tag{7.16}$$

We have chosen a case in which it is easy to find the element of area that is perpendicular to the current flow. This is necessary in order to obtain i from jA as we did in Eq. (7.15). If the current is not perpendicular to the cross section chosen, we must use $i = \mathbf{j}\cdot\mathbf{A}$, which greatly complicates the problem.

As we shall see in Sec. 7.9, the resistivity of metals varies with temperature and purity and is affected by imperfections in the crystal structure. In general, the resistivity in metals increases with increasing temperature, owing to the reduction in the length of electron paths between collisions with the lattice. In the case of semiconductors, discussed in Sec. 11.3, the resistance generally decreases with increasing temperature. This results from the higher density of charge carriers in semiconductors at higher temperatures.

Some materials and devices do not obey Ohm's law. For example, in a gas discharge tube (see Chap. 13) and in a diode vacuum tube (Sec. 11.2) or a crystal diode (Sec. 11.3), the current is *not* proportional to the applied voltage. In such cases the simple model of a constant number of conduction electrons undergoing collisions as they drift under the influence of an external field is not appropriate.

7.4 Energy Dissipation in a Resistance

When current passes through a resistance, electric energy is converted to thermal energy. When the potential difference between the two terminals of a resistance R is V, an amount of work $V\,dq$ is done by the electric field in moving a (positive) charge dq from the higher to the lower potential, or

$$dW = V\,dq$$

The rate at which work is done, or power, is

$$P = \frac{dW}{dt} = \frac{V\,dq}{dt} = Vi \tag{7.17}$$

From Eq. (7.7), $V = iR$, so

$$P = i^2R \qquad \text{joules/sec} = \text{watts} \tag{7.18}$$

or

$$P = \frac{V^2}{R} \tag{7.19}$$

These last two expressions are of course applicable only where Ohm's law applies.

The energy dissipated in the resistance goes into heat. The process is one in which electrons are accelerated by the applied field but continually lose their excess energy by collisions with the lattice of ions. This heats up the lattice by increasing the energy of thermal vibrations. Heating effects can be converted to thermal units through the relationship, one calorie is equal to 4.186 joules.

7.5 Electromotive Force, Internal Resistance

Since energy is dissipated when current passes through a resistance, as shown in the last section, any circuit in which there is a current and which involves dissipative elements such as resistances requires a source of energy. The energy source causing current flow in a circuit *cannot* be an electrostatic field. This follows from the fact that the change in electrostatic potential around any closed path is zero $\left(\oint \mathbf{E} \cdot d\mathbf{l} = 0\right)$. Thus an electrostatic field cannot give energy to charges that move completely around a circuit. The source of energy in a circuit may be chemical, thermal, or mechanical, or it may result from a changing magnetic field, as will be discussed in Chap. 8. In many cases the energy source is localized, as in a *chemical cell* or *battery*, but there are also situations in which the source of energy extends over the entire circuit. A number of particular sources will be discussed later, but here we develop a generalized method for describing all sources of energy causing current in closed circuits.

A source of energy is characterized by its *electromotive force* (emf) $\mathcal{E}$. This is defined by the equation

$$P = \frac{dW}{dt} = \mathcal{E}i \tag{7.20}$$

The power P is the rate at which energy from a nonelectrostatic source is converted to electric energy. This rate depends on the emf and the current i in the circuit. Letting $i = dq/dt$, Eq. (7.20) may be written as

$$\mathcal{E} = \frac{dW}{dq} \qquad \text{joules/coulomb, or volts} \tag{7.21}$$

Thus the emf is the work per unit charge done by the source. The unit of emf in mks or practical units is the volt. Because the unit of emf is the same as the unit of electrostatic potential difference, emf is often confused with electrostatic potential. The important difference between the two is that while an electrostatic potential difference can do work on a charge when the charge is moved from one point to another, only a source of emf can provide energy to cause current around a *closed* circuit.

In a steady state, the rate at which energy is put into an electric circuit is equal to the rate of dissipation of electric energy in the circuit. Thus Eq. (7.20) gives not only the rate of production of electric energy but also the rate of electrical dissipation in the circuit. If the losses in the circuit are all resistive and the total resistance in the circuit is R, it follows that

$$\mathcal{E}i = i^2R \tag{7.22}$$

Here we are assuming a simple circuit with only one path for the current (Fig. 7.7). The resistance R includes all the dissipative elements in the entire circuit.

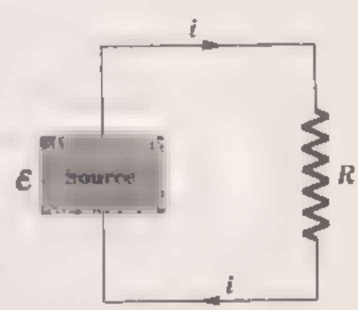


***Fig. 7.7** A source of electromotive force that produces a current in a closed circuit.*

In cases where the source of emf is localized, as in a battery, the losses in the circuit may be divided into losses *within* the source

of emf and losses in the external circuit. Figure 7.8 shows the way in which resistive elements can be divided into the *internal resistance* R_i and resistance in the external circuit, R. Upon dividing

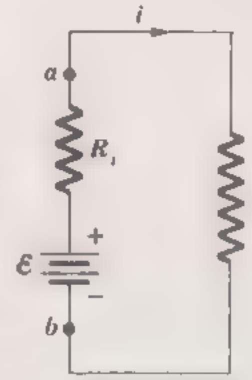


***Fig. 7.8** Schematic diagram showing internal resistance of a source of emf. Source shown is a chemical cell or battery. Battery terminals are at a and b.*

Eq. (7.22) by i and substituting $(R + R_i)$ for the total circuit resistance, we find

$$\mathcal{E} = i(R + R_i) \qquad \text{or} \qquad \mathcal{E} - iR_i = iR \tag{7.23}$$

The potential difference across the external circuit is given by $V = iR$. This is also the potential difference across the terminals of the source. We may then write

$$\mathcal{E} - iR_i = V \tag{7.24}$$

Figure 7.9 shows a plot of the potential difference across the terminals of an energy source as a function of the current through the

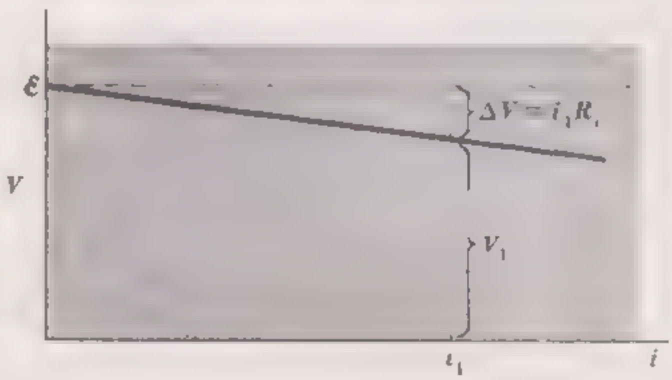


***Fig. 7.9** Output voltage of a source of emf versus current. At a current i_1 there is a resistive voltage drop $\Delta V = i_1 R_i$ within the source, leaving a potential difference V_1 between electrodes.*

source. The emf of the source equals the potential difference between its terminals when no current flows. When the internal resistance of the source is negligible compared with the resistance in the external source, we may write the approximate equation

$$\mathcal{E} = V \tag{7.25}$$

In the case of localized sources it is useful to look at a circuit from another point of view, as suggested by the diagram in Fig. 7.10. It may be considered that the nonelectrostatic source of energy causes a separation of positive and negative charge within the source. The separated charges produce an electrostatic field within

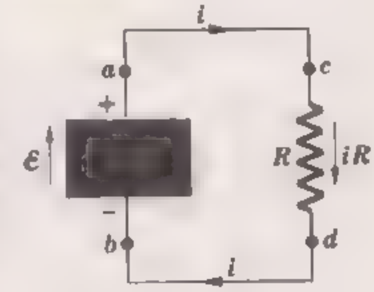


***Fig. 7.10** In this circuit the source of emf produces a potential difference $V = ℰ - iR_i$ between terminals a and b. The same potential difference is across the load cd, causing a current $i = V/R$ in the load.*

the source, and work must be done by the energy source to move charges against this field. A potential difference $ℰ - iR_i = V$ is thus maintained between the source terminals a and b. When wires of low resistance connect the terminals of the source to the load, this same potential difference is maintained across the load terminals c and d. From this point of view, the source of emf provides the electrostatic field that causes current in the load.

7.6 *Circuits, Kirchhoff's Rules*

We now discuss some general properties of electric circuits. Figure 7.10 is an example of a typical circuit containing a source of emf and a dissipative resistance. The usual problem is to calculate the current in the circuit and the voltage across each element. We have already developed Eq. (7.23), which gives the solution when we know the values of emf and resistances of the elements in the circuit. This equation holds around any *closed* circuit. In the general case where there is more than one source of emf and more than one resistance in the circuit, it may be rewritten as

$$\Sigma ℰ - \Sigma iR = 0 \tag{7.26}$$

The first term represents the emf between the terminals of the source or sources, and the second term is the sum of the voltage drops across the resistances in the circuit, including internal resistance in the source. The emf is positive in the direction in

which positive charge is displaced, and the voltage drop iR is positive in the direction of the current.

Often this equation is rewritten so as to include the effect of the internal resistance on the left-hand side, writing $V = \mathcal{E} - iR_i$ for the potential difference across the source when current flows. This gives

$$\Sigma V = \Sigma iR \tag{7.27}$$

where R now refers only to external resistances around the circuit. This is essentially the same as Eq. (7.26).

A second important rule comes from the idea of charge conservation, that is, that charge is neither created nor destroyed. It follows that if there is no piling up of charge at any point in a circuit, the rate at which charge comes to that point must be equal to the rate at which charge leaves the point. Thus the current to the point must equal the current away from the point, or at any point on the circuit,

$$\Sigma i = 0 \tag{7.28}$$

In this terminology, current coming into a point is given one sign, and current leaving the point is given the opposite sign. When applied to a simple circuit containing no branch paths, such as in Fig. 7.10, this rule says that the current has the same value at

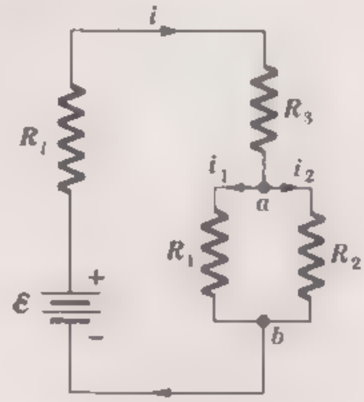


Fig. 7.11 ***Circuit with divided current path.***

every point in the circuit, which we have already tacitly assumed in writing Eq. (7.23).

In a more complicated circuit, such as shown in Fig. 7.11, we apply this same rule at the junctions a and b. It tells us that the current i coming into the junction equals the sum of the currents i_1 and i_2 through the parallel resistances R_1 and R_2.

In circuits like that of Fig. 7.11, we can obtain currents and voltages by applying Eqs. (7.27) and (7.28) and by using the rules for series and parallel combinations of resistances already developed. Many complicated circuits can be broken down into simple series and parallel combinations of resistances, allowing simple solution.

We give a formal solution of the circuit of Fig. 7.11, assuming $\mathcal{E}$ and all the resistances are known. We can write the circuit equation

$$\mathcal{E} = iR_i + iR_3 + V_R$$

where V_R is the voltage across R_1 or R_2. Also,

$$V_R = i_1R_1 = i_2R_2$$

and

$$i_1 + i_2 = i$$

Finally, the equivalent resistance of R_1 and R_2 in parallel is given by

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \qquad \text{or} \qquad R = \frac{R_1R_2}{R_1 + R_2}$$

The circuit equation becomes

$$\mathcal{E} = iR_i + iR_3 + i\frac{R_1R_2}{R_1 + R_2}$$

which we can now solve for i. The values of i_1 and i_2 are readily obtained once i is known.

In some circuits the resistance cannot be handled in the simple series or parallel fashion. Such a circuit is shown in Fig. 7.12. This circuit will be used to describe a method for finding the current in each branch, given a knowledge of the emf $\mathcal{E}$ and of all resistances. We use i_1 for the current through R_1, and so forth. The rules for solving such a problem are called *Kirchhoff's rules*, which are actually simply the circuit and current equations we have just developed,

$$\Sigma\mathcal{E} - \Sigma iR = 0 \qquad \text{and} \qquad \Sigma i = 0 \qquad (7.26,\ 7.28)$$

The first equation applies to any closed loop, and the second is to be applied at junctions. The application of these rules is greatly simplified by establishing a definite procedure. We begin by

making assumptions regarding the direction of current that we call positive in each branch. It is not necessary that our choice be correct, since if a current comes out negative it merely means that it flows the opposite way. Correctly assigned current directions will be positive.

We next prepare to apply the circuit equation to several of the loops. (It is not necessary to consider *all* possible loops. We need only a total number of equations equal to the number of unknown currents.) For each loop we decide on the direction we choose to go around the loop, as illustrated by the paths A, B, and C in Fig. 7.12. Care must be taken in choosing the signs of the emfs and iR drops as we proceed around the loops. Just as in the case of the simple circuit of Fig. 7.10, $\mathcal{E}$ is taken positive if we move through the source from negative to positive. (As we have chosen the direction of path C in Fig. 7.12, $\mathcal{E}$ is negative.) The voltage drops iR through the resistances are substituted into Eq. (7.27) as positive quantities if we move in the same direction as the assigned current, and as negative quantities if we travel in a direction opposite to the current. To illustrate these ideas, we list the equations for circuits A, B, and C:

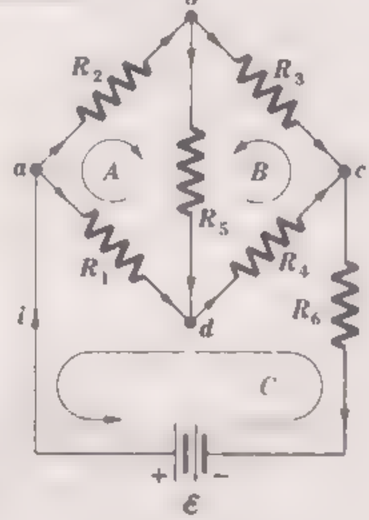


***Fig. 7.12** Use of Kirchhoff's rules for current calculations.*

Loop	*Equation*
A	$-i_1R_1 + i_2R_2 + i_5R_5 = 0$
B	$i_4R_4 - i_3R_3 + i_5R_5 = 0$
C	$-\mathcal{E} - i_6R_6 - i_4R_4 - i_1R_1 = 0$

We next write the junction equations according to Eq. (7.28):

Junction	*Equation*
a	$i_6 = i_1 + i_2$
b	$i_2 = i_3 + i_5$
c	$i_6 = i_3 + i_4$
d	$i_4 = i_1 + i_5$

Once the equations are written down, the problem of solution is entirely algebraic. It usually pays to replace the symbols for resistances by their actual values since this simplifies the algebra.

The procedure given is only one of several equivalent methods. However, all methods depend on the application of the same two principles of conservation of energy and conservation of charge.

7.7 Current and Voltage Measurement, the Galvanometer

We have discussed electric currents at some length without giving any details as to how currents can be measured in the laboratory. We now make up this deficiency and start with a discussion of the galvanometer, the most used device for measuring both current and, indirectly, potential difference.

As mentioned briefly in Chap. 6, a torque is exerted on a current-carrying loop placed in a magnetic field as shown in Fig. 7.13. In the galvanometer, a loop carrying the current to be

Fig. 7.13 *Galvanometer coil in a magnetic induction field.*

measured is placed in the magnetic field of a permanent magnet. A spring acts to hold the plane of the loop parallel to the magnetic field, and current in the loop causes the loop to twist toward the perpendicular to the field against the torque exerted by the spring. Motion of the loop actuates a pointer or, in the most sensitive instruments, moves a mirror so as to cause a light beam to move over a scale.

The sensitivity of a galvanometer can be determined through the application of our earlier equation,

$$\tau = NiAB \sin \theta \tag{6.21}$$

giving the torque on a coil of N turns, area A, carrying a current i in a magnetic induction field B. Here we have used θ for the angle between the normal to the plane of the coil and the field direction. The angular deflection θ for a given current will depend on the restoring torque provided by the coil suspension. We characterize this by the torque constant k, where

$$\tau = k\theta \tag{7.29}$$

When the plane of the coil is parallel to the field, $\theta = 90°$, or $\sin \theta \sim 1$, and we may equate (6.21) and (7.29) to give

$$i = \frac{k}{NAB}\theta = K'\theta \tag{7.30}$$

where K' is called the galvanometer constant. Thus the deflection angle of the galvanometer is proportional to the current through it. In practice this equation is made applicable over a wide angle by shaping the magnetic field so that the lines of B remain nearly parallel to the coil as it turns, as shown in the sketches in Fig 7.14.

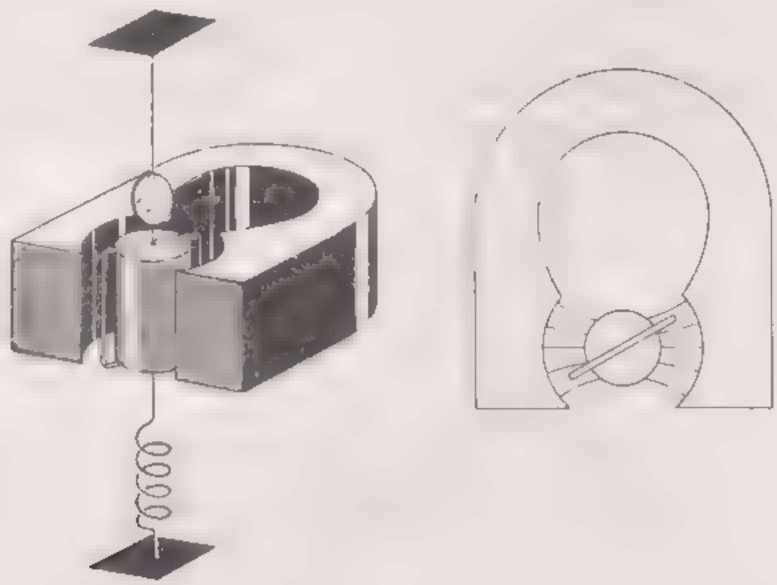


Fig. 7.14 Galvanometer coil placed between magnet poles. Soft iron core produces radial field.

A soft iron cylinder is often placed at the center of the coil, which together with the specially shaped poles of the permanent magnet causes the magnetic field lines to be radial.

A common practice with sensitive galvanometers is to quote sensitivity in terms of the current necessary to cause a 1-mm deflection of a light beam reflected by the galvanometer mirror on a scale 1 m away. Since the light beam is deflected by an angle 2θ, we have $d/L = \tan 2\theta \simeq 2\theta$, where L is the distance from galvanometer to scale. Substitution in Eq. (7.30) gives

$$i = K' \frac{d}{2L} = \frac{K}{2{,}000} d = Kd \tag{7.31}$$

where d is in millimeters. K is called the *figure of merit* of the galvanometer. Currents as small as 10^{-11} amp can be measured on the most sensitive galvanometers.

When the galvanometer is used as a current-measuring instrument, it is desirable that its resistance be very low compared with other elements in the circuit to which it is connected. This is necessary to avoid perturbing the circuit. A simple example is shown in Fig. 7.15. Here we might wish to determine the resistance R by

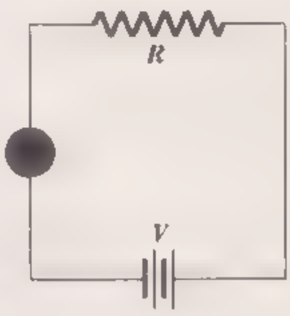


Fig. 7.15 Galvanometer being used to measure current.

measuring the current flow through it when it is connected to a known voltage source V. We should like to have the resistance of the galvanometer much less than R so that all but a negligible fraction of the voltage V appears across R. Galvanometers for use in ammeters are wound with a few turns of heavy wire to keep the resistance as low as possible.

The maximum sensitivity of an ammeter is fixed by the sensitivity of the galvanometer movement of which it is made. However, it is often required to reduce the sensitivity to a lower value. This is done by providing a *shunt* or parallel resistance, as shown in Fig. 7.16. Here the resistance of the galvanometer movement is given by R_g, and the shunt resistance is R_p. Application of Ohm's law to this circuit gives the value of R_p necessary to reduce

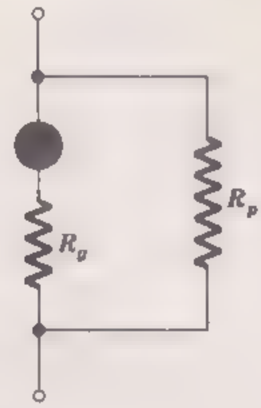


***Fig. 7.16** Shunt resistance R_p used to adjust sensitivity of galvanometer used as an ammeter.*

the sensitivity to any desired value. Such a shunt also reduces the total resistance of the ammeter.

When a galvanometer is to be used as a voltmeter, as in Fig. 7.17, it is necessary that its resistance be much larger than

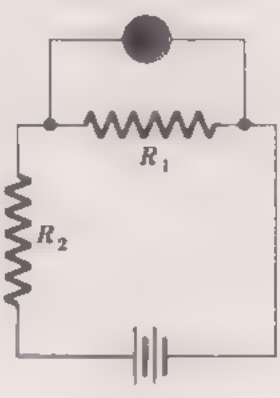


***Fig. 7.17** Galvanometer used as a voltmeter, to measure voltage across R_1.*

any resistance across which it is to be placed. Otherwise the voltmeter will alter the current flow in the circuit, which is usually undesirable. Galvanometers to be used as voltmeters are wound with many turns of fine wire to give high sensitivity and high resistance. Usually there is sensitivity to spare, and the total resistance of the voltmeter is increased and its sensitivity is reduced to the desired value by the addition of the appropriate series resistance R_s as shown in Fig. 7.18. A simple calculation

***Fig. 7.18** Series resistance R_s used to adjust sensitivity of galvanometer used as voltmeter.*

again allows the determination of the appropriate value of R_s for a given sensitivity.

7.8 The Potentiometer

The most common circuit used for measuring potentials is the potentiometer This is actually a voltage comparison device that uses a galvanometer as a null indicator. A simple form of this circuit is shown in Fig. 7.19. It consists of a voltage source V, con-

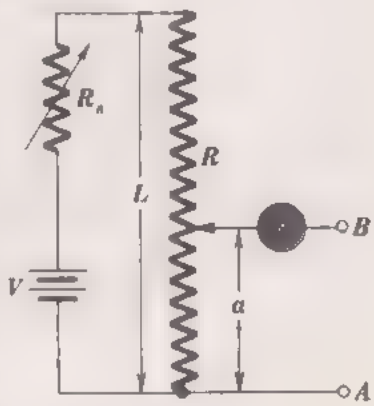


Fig. 7.19 Potentiometer circuit.

nected through an adjustable resistance R_s to a resistance wire R. The voltage across the terminals AB is adjustable by moving the sliding contact to B. The special feature of the resistance R is that it is made of uniform wire so that the voltage across AB is related to the total voltage across R by the ratio of the length a to L. That is,

$$\frac{V_{AB}}{V_R} = \frac{a}{L}$$

This follows at once from Ohm's law, supposing that the wire is uniform in resistivity and cross section.

In use, the potentiometer is first adjusted and calibrated by measuring the voltage of a known source such as a standard cell. That is, terminal B is arbitrarily placed at some convenient position (for example, the distance a may be calibrated in volts, in which case B is set at the voltage of the known source), and the adjusting resistor R_s is moved until the galvanometer reads zero. This in effect establishes a standard current in the potentiometer. An unknown potential is then measured by placing it across AB and adjusting the distance a until the galvanometer again reads zero. A simple Ohm's-law calculation then gives the potential of the unknown voltage. When the potentiometer is balanced, no current is taken from the voltage source being measured. In actual practice, the galvanometer is shorted by a shunting resistance to de-

crease its sensitivity until balance is almost perfect. This prevents damage to the galvanometer from overloading. In a similar way the standard cell usually associated with the potentiometer is protected from heavy current drain, which will damage it, by a large series resistance that is shorted out after almost perfect balance has been reached. In most potentiometers the single slide wire is replaced by a series of taps and a slide wire that allows for a similar accurate adjustment and measurement of the effective ratio a/L. The term *potentiometer* usually refers to the entire apparatus, including battery, standard cell, and adjustable resistance tap used for measuring unknown potentials. However, any adjustable tap on a constant resistance that allows a variable voltage to be obtained is also called a potentiometer. Figure 7.20 shows the contrast

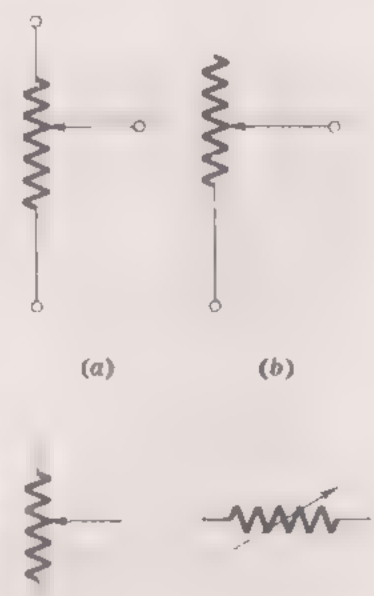


Fig. 7.20 *Variable resistance connected as (a) potentiometer (three terminals) and (b) rheostat or variable resistor (two terminals). Symbols used as shown.*

between a variable resistance, or *rheostat*, and a potentiometer as used in electronic circuits. The same device may be used either as a potentiometer or as a rheostat. The only difference is that in a potentiometer current flows through the entire resistance element and a variable voltage is tapped off, making it a three-terminal device, whereas a rheostat is a two-terminal device in which current flows through a variable amount of the resistance element.

7.9 Wheatstone's Bridge

Another circuit used for measuring resistances is Wheatstone's bridge (invented in 1833, probably by a man named Christie),

shown in Fig. 7.21. This is also a null instrument using a galvanometer to indicate "balance." R_1 is here the unknown resistance and R_2, R_3, R_4 are known resistances, one or more of which are

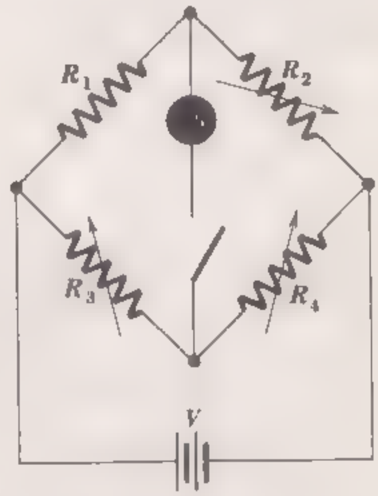


***Fig. 7.21** Wheatstone's bridge, allowing measurement by null method of an unknown resistance R_1 in terms of three other known resistances (R_2, R_3, R_4).*

adjustable. No current flows when the switch in the galvanometer circuit is closed, provided that the resistances are related so that

$$\frac{R_1}{R_3} = \frac{R_2}{R_4} \tag{7.32}$$

This allows the unknown resistance to be calculated. This circuit is one example of many *bridge* circuits that allow null measurements to be made, giving high accuracy. This method does not require knowledge of or strict constancy of the voltage source V.

7.10 Variations in Resistivity of Metals

A few remarks about the variation in resistivity among different kinds of metals under different conditions are in order. Table 7.1 gives resistivities of a few typical metals at 20°C.

Table 7.1

Metal	*Resistivity ρ, ohm-m*
Cu	1.7×10^{-8}
Al	2.8×10^{-8}
Ag	1.6×10^{-8}
Au	2.4×10^{-8}
Brass	$\sim 7 \times 10^{-8}$

In Sec. 7.3 we developed a simple classical theory of conductivity A detailed understanding of the behavior of conduction electrons in metals involves quantum-mechanical considerations that we shall not present here. However, we may still use the classical expression derived earlier,

$$\rho = \frac{m}{Ne^2\tau} \tag{7.10}$$

realizing that the quantities N, τ, and possibly even m will have to be evaluated by quantum-mechanical methods. Thus we may use Eq. (7.10) to get a qualitative idea of how the resistivity varies under various conditions.

For a given metal the main source of variation of resistivity with changing conditions is in the variation of τ, a measure of the mean time between electron collisions with the lattice of ions that make up the metal. Thus anything that affects the likelihood of these collisions influences the resistivity. Because of the quantum- or wave-mechanical behavior of the conduction electrons in metals, the ideas of wave mechanics are necessary to describe the important effects. One very important result is that conduction electrons can move through a perfectly periodic lattice containing uniformly spaced identical ions with essentially no collisions. This is in marked contrast with the random collisions to be expected on a classical picture and results from the wave nature of electrons. However, no real metal has an ideally perfect lattice, and in consequence there are collisions that seriously limit the mean free path or time between collisions. For example, the purity of a metal is not of importance in determining its resistivity at room temperature. There is a relatively small difference between ρ for pure copper and for brass, which, being made of a random mixture of copper and zinc atoms, is far from a perfectly periodic lattice. The reason for this small effect lies in the fact that at room temperature, thermal vibrations, even in very pure copper, prevent the lattice from being truly evenly spaced or periodic, so that τ is already seriously limited by the displacements of ions by thermal vibrations, and the additional effects due to random mixing of unlike ions are not important. In a similar way, strains and other imperfections introduced into the metal make rather small changes in the room-temperature resistivity.

The full implications of the wave-mechanical model are realized, however, when the metal is at extremely low temperature. In copper at 4°K, the approximate temperature of boiling liquid helium at atmospheric pressure, the effects of thermal vibrations are reduced to insignificance compared with other sources of lattice collisions. This is shown by Table 7.2, which gives some typical results for the ratio between resistivity at room temperature (300°K) and at 4°K for various samples of copper.

Table 7.2

Sample	*Resistivity ratio*
Very pure unstrained copper	**2,000–7,000**
Ordinary unstrained copper	**50–100**
Brass (copper + zinc)	**4**

The very large ratio in the purest metal signifies that at room temperature most of the collisions that limit τ are due to thermal displacement of the otherwise periodically spaced ions. The smaller ratio in less pure copper shows that even at room temperature, many of the collisions are due to impurity ions and to imperfections in the crystal structure These effects persist at low temperatures and therefore limit the ratio of resistivities to relatively low values. In the case of brass, impurity effects due to the two kinds of ions dominate, and the resistivity ratio is very small.

Hall-effect measurements show that the density of charge carriers is essentially temperature-independent in metals; this makes these measured variations of resistivity with temperature very hard to understand on a classical basis.

In view of the dependence of resistivity on quantum-mechanical details, not all of which are easily determined, we shall not comment further. We shall, however, explain the usual notation that is used to describe the temperature variation of resistance in the room-temperature range. Over a reasonable range (from 0 to, say, 200°C) resistivities of most metals are found to increase linearly with temperature. The resistivity in this linear range at a temperature T can then be written as

$$\rho_T = \rho_{20°\mathrm{C}}(1 + \alpha\, \Delta T) \tag{7.33}$$

where ΔT is the difference between 20°C and the temperature T.

The constant α is the fractional change in resistivity per degree. For most metals α lies in the range of 3.5 to 4 parts in 1,000 per degree centigrade. In certain alloys that have been developed empirically, α is as low as 1 part in 10^5 per degree over an appreciable temperature range. Such alloys are particularly useful for making fixed resistances where variation with temperature is undesirable.

7.11 Electromotive Force of a Chemical Cell

One of the earliest practical sources of electric energy, and one that is still of great importance, is the chemical cell or battery. There are many different types, but they all depend on the same principles. Most students will have studied this problem already in a chemistry course, so the discussion here will be brief.

In order to understand the production of an emf by chemical forces, let us examine the experimental facts. If a piece of metal, say copper, is placed in a dilute acid solution, there is a tendency for Cu^{++} ions to leave the metal surface and go into solution. This tendency is called *solution pressure* and results from the attraction of water molecules for the metal ions that overcomes the attraction of the metal surface for the ion. Each ion that goes into solution leaves its electronic charge behind, so the potential of the metal with respect to the solution becomes increasingly negative. The metal continues to dissolve until an equilibrium state is reached. This occurs when the rate at which ions escape from the metal is equal to the rate at which they come back to the surface from the solution. Since the rate at which ions come back and stick to the metal surface depends on the number that strike the surface per second, the equilibrium depends on the concentration of ions in solution (and also on the temperature, which affects their velocity). The equilibrium is also affected by the potential difference that has been reached as a result of the electronic charges left behind by the escaping ions. For a given concentration of ions of the metal in solution, each kind of metal comes to a characteristic negative potential upon reaching equilibrium. If we place two electrodes, say Cu and Zn, in a solution, the Zn becomes more negative than the Cu (owing to its greater solution pressure). In this situation, though no net flow of Cu^{++} ions to the Cu is occurring, the more negative Zn electrode attracts Cu^{++} ions,

which are deposited as Cu atoms and take up some of the net negative charge on the Zn electrode. The equilibrium potential on the Zn electrode is maintained if, on the average, for each Cu^{++} ion captured, a Zn^{++} escapes. Thus there is a plating of Cu on the Zn, and Zn^{++} ions will come into the solution. The equilibrium situation involves a very low Cu^{++} ion concentration and a high Zn^{++} ion concentration.

Suppose the two electrodes are connected together externally, say through a voltmeter. Electrons flow from the more negative Zn to the Cu electrode. This decreases the negative potential of the Zn electrode, allowing more Zn^{++} ions to go into solution. The Cu electrode tends to become more negative and thus allows more Cu^{++} to come back from solution. As long as this process can continue, the chemical action continues to maintain the electrode potential difference and we have a useful battery.

One factor that influences the measured emf of the battery is the *contact potential difference* between the two metals. Two dissimilar metals when placed in contact assume a characteristic difference in potential owing to the escape of electrons from one metal to the other. We discuss this in Sec. 7.12. This contact potential affects the net voltage measured across the electrodes. By measuring the emf between pairs of electrodes, a table of emfs can be constructed. Table 7.3 shows such a listing, where potentials

Table 7.3 Electromotive Force Series for Some Elements

(Relative to the hydrogen electrode)

Element	*Ion*	*Potential at 25°C, volts*
Li	Li^{+}	**−2.959**
Na	Na^{+}	**−2.715**
Zn	Zn^{++}	**−0.762**
Fe	Fe^{++}	**−0.44**
Cd	Cd^{++}	**−0.402**
Sn	Sn^{3+}	**−0.336**
Pb	Pb^{++}	**−0.12**
Pt, H_2	H^{+}	**0.000**
Cu	Cu^{++}	**+0.345**
Hg	Hg^{++}	**+0.799**
Ag	Ag^{+}	**+0.798**

are given relative to a hydrogen electrode (a Pt electrode over which H_2 gas is bubbling). A standard ion concentration of 1 mole/liter is used for these measurements. The emf between any two kinds of electrodes is the algebraic difference between the emfs listed.

We define the emf of a battery as the potential difference between its terminals when no current is allowed to flow. There is in general a reduction in this potential difference when current flows, because the ions carrying the current must be given an extra drift velocity and the energy for this must come ultimately from the chemical forces. This wasted energy shows up in the heating of the cell when current passes through it and is the cause of internal resistance in the cell.

For practical purposes we may classify batteries into those which wear out by using up the available chemical energy and those in which a reverse chemical reaction caused by forcing current backward through the cell renews the cell to its original state. Elementary chemistry books give examples of both types. Of particular importance in precise electrical measurements is the standard cell, which is constructed to give a very accurately known potential as long as no current is allowed to flow. Such cells must be used in potentiometer or other circuits that allow voltage measurement without drawing appreciable current.

7.12 Contact Potential and Thermal Electromotive Force

In Sec. 7.11 we mentioned the difference in potential between two dissimilar metals in contact. In order to understand this and some related phenomena, we must look at the behavior of conduction electrons from a quantum-mechanical point of view. One of the most drastic quantum effects is on the distribution of energies or velocities of conduction electrons. On a classical basis, each conduction electron is expected to have a mean kinetic energy of $\frac{3}{2}(kT)$, where k is the Boltzmann constant and T the absolute temperature. Classically there is no prohibition on many electrons having the same energy at the same time. In contrast, it is a central result of quantum mechanics that for electrons that can move within a metal, only certain energies are allowed and each energy level can be occupied by no more than two electrons (in pairs having opposite spin orientation). The consequence of this is that

in order to accommodate the large number of conduction electrons in a metal (one or more per atom) by successively filling the allowed energy levels with pairs of electrons, the energy of the highest levels that must be occupied is very high compared with the classical expectation. The energy at the top of this filled band of energies is equivalent to the mean energy of a classical electron at a temperature of over 10,000°K. Calculations that we shall not go into show that the distribution of conduction-electron energies is as given in Fig. 7.22. The distribution curve is shown at 0°K and at a high

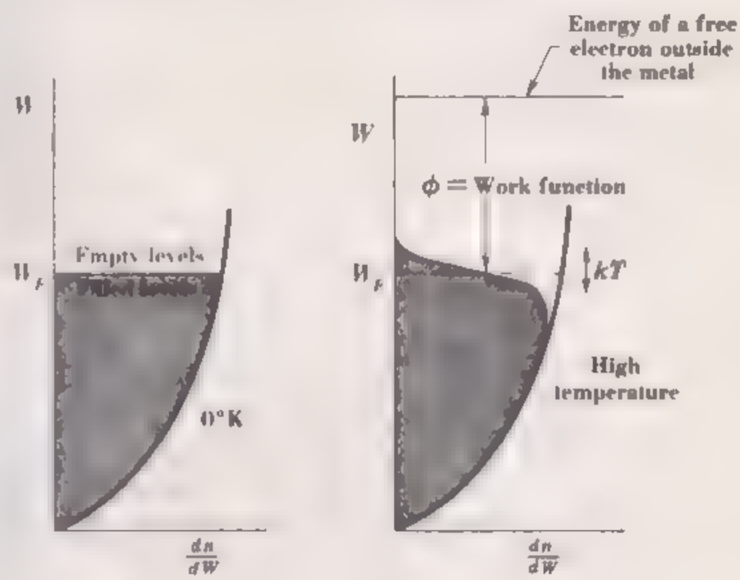
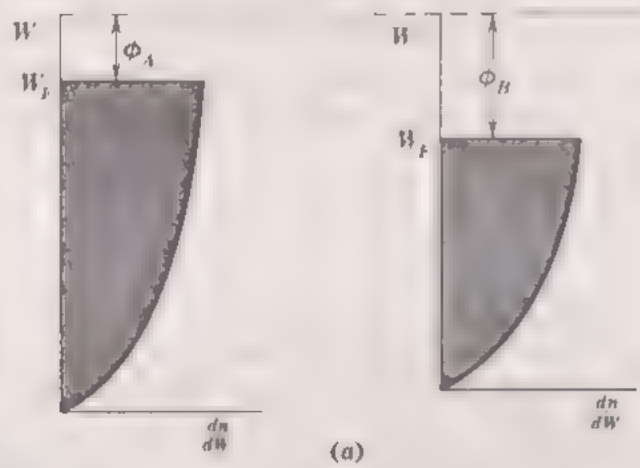


***Fig. 7.22** Fermi distribution of conduction-electron energy levels in a metal. dn/dW is the number of available energy levels per unit volume, per unit energy. At 0°K all levels are filled up to W_F, the Fermi level. At high temperature some electrons near the top of the distribution are thermally promoted to higher levels above W_F, leaving corresponding vacancies among the upper levels that are filled at 0°K.*

temperature. These curves show the way in which the allowed energies are filled with electrons. The quantity dn/dW is the number of allowed energy levels in a given energy range dW The shaded part of the plot shows the levels occupied by electrons. At very low temperatures all levels are filled up to some maximum value called the Fermi level W_F. At high temperatures some electrons near the top of the distribution are excited to higher levels above W_F, leaving behind some unoccupied levels.

The energy necessary to remove an electron from the top of the energy distribution and allow it to escape from the metal is called the *work function*. Differences in the work function among metals produce the contact potential between dissimilar metals. This we can show by following the sequence of events that occur when two metals are brought into contact. Figure 7.23*a* shows energy diagrams of two metals A and B, which have not yet been placed in contact. The two energy coordinates have been adjusted

to give the same energy for electrons that can just escape. We call this minimum escape energy W_∞. Because of the difference of the work functions ϕ_A and ϕ_B, the Fermi energies W_F are at different levels on the energy scale. When the two metals are brought into close contact, electrons at the top of the distributions are



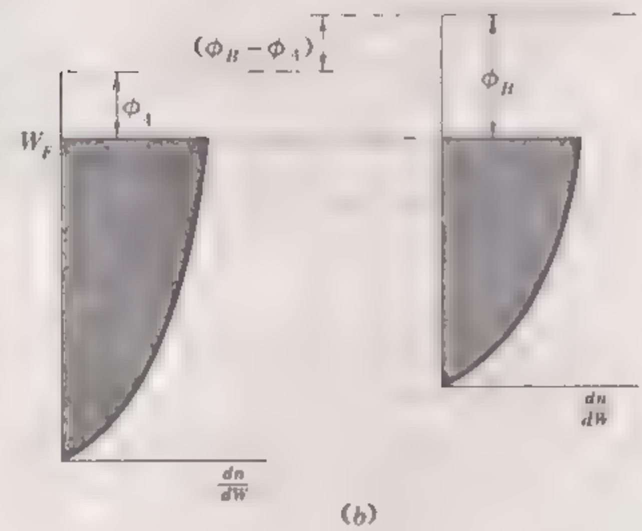


***Fig. 7.23** (a) Schematic representation of potentials of two uncharged metals, A and B. Absolute values of energies are adjusted to give same energy for electrons which have minimum energy to escape, W_∞. (b) Potentials of the two metals when in contact. Charge redistributes to bring Fermi levels together. Contact potential difference is $(\phi_B - \phi_A)$.*

free to move back and forth from one metal to the other. The net flow of electrons is from the higher to the lower Fermi level. The transfer of charge results in the respective raising and lowering of the potential of the two pieces of metal as shown in Fig. 7.23*b*. The net flow stops when the energy of the electrons at the top of the distribution is the same for each metal. Relative to the density of electrons in the conduction band, the number of electrons that must move from one metal to the other in order to shift the energies to an equilibrium position, by changing the electrostatic potential, is very small. Thus in the final state the limit of filled states is

sensibly unchanged relative to the metal. That is, the major effect is to cause the reference level of the energy scales of the two metals to slide into a relative position that equalizes the two Fermi levels. Thus, as is clear from the diagram, we expect the contact potential difference to be the difference between the work functions. Work functions are in the range of 1 to several electron volts.

If a complete circuit is made, containing several kinds of metal, the contact potentials at the junctions cancel, and there is no net emf around the circuit. However, since the contact potential varies with temperature, if the two junctions are at different temperatures there is a net emf equal to the difference between the two emfs. This is the well-known thermal emf used in thermocouples for temperature measurement. This effect was discovered by Seebeck in 1821. Usual values of the emf are in the range of 10^{-6} volt/°K. Thermocouples are not linear in their variation of emf with temperature and so must be empirically calibrated for accurate measurements The converse of the Seebeck effect is called the Peltier effect, which results in the cooling of a bimetal junction when the current passes in one direction and a heating for current in the opposite direction. In actual experiments, resistance or ohmic heating will be superposed on the heating or cooling from the Peltier effect.

The Thomson effect is closely associated with the Peltier and Seebeck effects. In a *single* metal, when one end is hotter than the other, the electrons from the hotter end tend to diffuse slightly faster than those from the cold end. This results in a small potential difference between the two ends, sufficient to counteract the net flow that would otherwise occur. The effect can be observed by passing a current through a conductor along which a thermal gradient is maintained. In addition to the i^2R heating, there will be an extra heating or cooling term depending on the relative directions of the thermal and electric currents.

These thermally induced emfs provide a means for converting heat energy directly into electric energy. A great deal of work is now going on to develop practical power sources using thermal emfs. However, the low efficiency of conversion has thus far prevented such sources from being of much practical importance.

PROBLEMS

7.1 The current density in a conductor of circular cross section of radius a varies with radius according to $j = j_0 r$. Find the total current.

7.2 A current flows from left to right in a conducting wire Is there any way in which the sign of the charge carriers can be determined?

7.3 A current of 10 amp flows through a wire of 1 mm^2 cross section. If the density of charge carriers in the wire is $10^{27}/m^3$, find the average drift velocity of the electrons.

7.4 The resistivity ρ of copper at room temperature is 1.7×10^{-8} ohm-m. If the density of mobile electrons is $10^{27}/m^3$, find the relaxation time τ for electrons in copper.

7.5 Find the resistance of a block of copper of length 20 cm and cross-section area 2 cm^2. The conductivity σ of copper is 0.59×10^8 ohm-m^{-1}.

7.6 Twelve resistors, each of resistance R, are joined to form the edges of a cube. Find the equivalent resistance between two diagonally opposite corners on a face of this cube.

7.7 A thin square sheet of uniform material is connected to two low-resistance conductors along opposite sides as shown in Fig. P7.7. If the sheet is 1 in. on each side, the resistance between the two conductors is 1 ohm. What would be the resistance if the same kind of arrangement were used with a sheet 2 in. on each side?

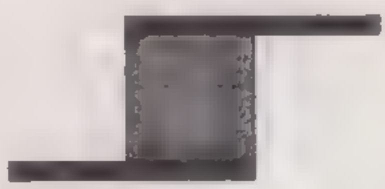


Fig. P7.7

7.8 *a* Calculate the resistance between points A and B in Fig. P7.8. $R_1 = 2$ ohms, $R_2 = 3$ ohms, $R_3 = 4$ ohms, $R_4 = 6$ ohms, $R_5 = 5$ ohms.

b Calculate the power dissipation in each resistor if the potential difference between A and B is held at 10 volts.

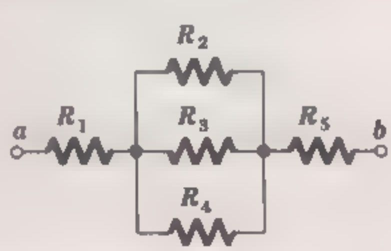


Fig. P7.8

7.9 A block of material 10 cm long and 2×1 cm in cross section has a resistance between ends of 10^{-4} ohm. What will be its resistance if

it is deformed so as to be only 5 cm long and of uniform cross section, assuming no change in its resistivity?

7.10 In Fig. P7.10,

a Find the equivalent resistance of the network between points *a* and *b* in terms of R_1, R_2, R.

b Find the numerical value of the equivalent resistance if $R_1 = 4$, $R_2 = 2$, $R = 1$.

c Compare (*b*) with the equivalent resistance when R is removed.

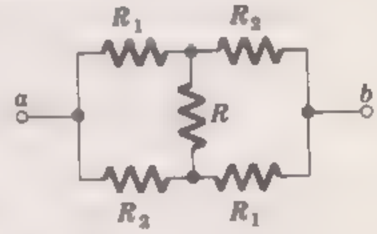


Fig. P7.10

7.11 Find the effective resistance (resistance between *a* and *b*) of an infinitely long ladder of resistors, as shown in Fig. P7.11, each having resistance R.

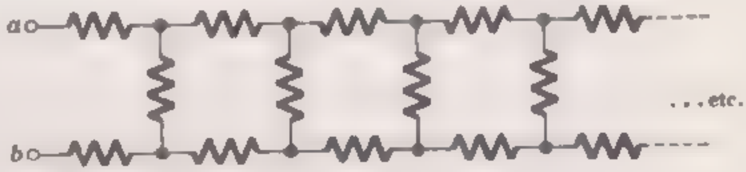


Fig. P7.11

7.12 A source of emf causes a constant current i to flow through a resistance R. What is the work done during the time a total charge Q passes through the resistor? What is the rate at which work is being done? What is the rate of heat generation in calories per second in the resistor?

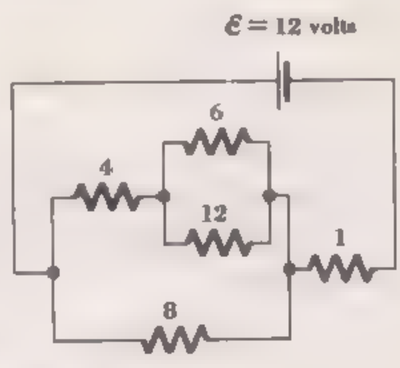


Fig. P7.13

7.13 In the circuit shown in Fig. P7.13 the internal resistance of the battery (not shown) is 1 ohm. The resistance of each resistor in ohms is given in the diagram. Find:

a The current in the battery

b The current in each resistor

c The power dissipation in each resistor and in the internal resistance of the battery

d The power generated by the chemical forces in the battery

e The potential difference across the terminals of the battery

7.14 The internal resistance of a dry cell gradually increases with age, even though the cell is not used. The emf, however, remains fairly constant at about 1.5 volts. Dry cells are often tested for age at the time of purchase by connecting an ammeter directly across the terminals of the cell and reading the current. The resistance of the ammeter is so small that the cell is practically short-circuited.

a The short-circuit current of a fresh No. 6 dry cell is about 30 amp. Approximately what is the internal resistance?

b What is the internal resistance if the short-circuit current is only 10 amp?

7.15 An unknown emf $\mathcal{E}$ is to be measured by a potentiometer circuit, as shown in Fig. P7.15. Show that when the circuit is unbalanced, the current flowing through the galvanometer is given by

$$i_g = \frac{[R_2/(R_1 + R_2)]\,V - \mathcal{E}}{[R_1R_2/(R_1 + R_2)] + R_g}$$

The internal resistance of the battery V and the resistance of the galvanometer are neglected.

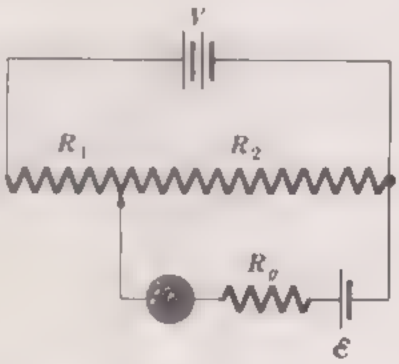


Fig. P7.15

7.16 Find the current in each branch of the circuit shown in Fig. P7.16. $V_1 = 5$ volts, $V_2 = 2$ volts, $R_1 = 3$ ohms, $R_2 = 2$ ohms, $R_3 = 4$ ohms.

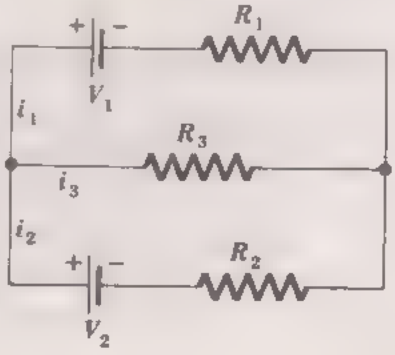


Fig. P7.16

7.17 *a* A galvanometer gives full scale deflection when $i = 1$ milliampere (ma). It is desired to make it into a voltmeter giving full scale deflection for 10 volts. The internal resistance of the galvanometer is 10 ohms. Tell how to do this. (Give diagram and necessary information.)

b If it is needed, instead, as an ammeter to read full scale for 10 ma, how can this be done? (Give diagram and necessary information.)

7.18 Figure P7.18 shows the internal wiring of a three-scale voltmeter whose binding posts are marked 3, 15, and 150 volts, respectively. The resistance of the moving-coil galvanometer used is 15 ohms, and a current of 1 ma in the coil causes it to deflect full scale. Find the resistances and the over-all resistance of the voltmeter on each of its ranges.

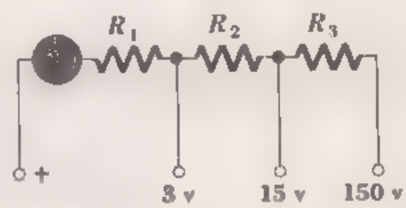


Fig. P7.18

7.19 The resistance of the moving-coil galvanometer in the ammeter shown in Fig. P7.19 is 25 ohms, and it deflects full scale with a current of 0.01 amp. Find the magnitude of the resistances required to make a multirange ammeter deflecting full scale with currents of 10, 1, and 0.1 amp, respectively.

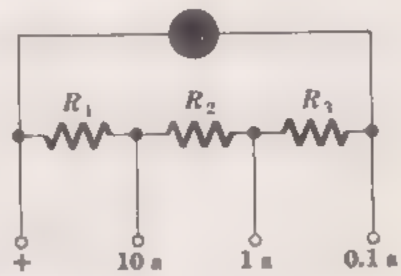


Fig. P7.19

7.20 In the circuit of Fig. P7.20 the internal resistance of the battery is R_i. Find the ratio of R_i/R_L to obtain maximum power dissipation in R_L, the load resistance.

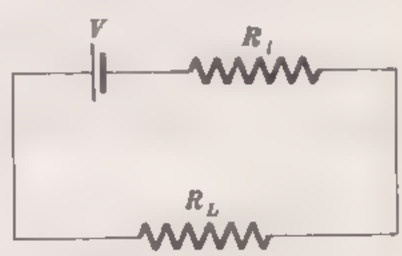


Fig. P7.20

EIGHT

Induced Electromotive Force, Inductance

8.1 Introduction

We have so far investigated the electric field due to static charges and the magnetic field of moving charges. In this chapter we add a third cornerstone to the theoretical structure by considering effects due to *changing* magnetic fields. We shall begin by looking in detail at some consequences of motion of circuits in a magnetic induction field, basing our conclusions on effects we have already studied. This leads us to a rephrasing of earlier ideas in terms of induced electromotive force. Finally, we show how both experiment and some general considerations lead naturally to Faraday's law of induction. This prepares for an examination of inductance as a circuit element in a-c circuits and is also crucial in the understanding of electromagnetic waves.

8.2 Motional Electromotive Force

In general, when a circuit of fixed shape is moved in a region containing a magnetic induction field, an emf is induced around the

circuit. We may calculate the magnitude of this effect by applying the rules for the magnetic force on a moving charge. Figure 8.1 shows a simple case in which a conducting rectangular loop is moving to the right in a field which is perpendicular to the plane of the loop and which varies only along the x direction. We now calculate the magnetic force on a charge q, at rest at a fixed position on the loop. Equation (6.25) gives this force as $F = qv_xB_z$, where F is upward in

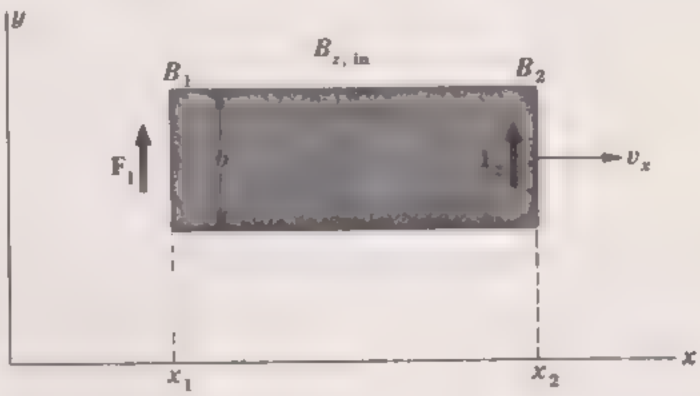


***Fig. 8.1** Motional emf. Motion in a magnetic induction field produces current in a conduction loop.*

the y direction for q positive, B_z is the magnetic induction field into the paper, and v_x is the velocity with which the loop is moved in the x direction. The magnetic force will act on charges on the wire and will do work on those which can move along the vertical sides of the loop. Only these sides need be considered since the force is perpendicular to the horizontal sides, so no magnetic work is involved in moving charges along them. The emf around the loop may be found by calculating the work per unit charge performed by the magnetic force in moving a charge around the loop. We consider counterclockwise forces as positive, in agreement with the earlier assignment of positive to counterclockwise loop current. As shown in Fig. 8.1, then, the upward magnetic force F_2 at the right-hand side of the loop is positive, while the upward magnetic force F_1 at the left-hand side is negative. The net magnetic force around the loop is thus $F_2 - F_1$. To clarify the situation, let us assume that B_2 at the right-hand side of the loop is greater than B_1 on the left. The net magnetic force is then counterclockwise or positive.

The work done by the magnetic forces in carrying a charge q around the loop is given by

$$W = \oint \mathbf{F}\cdot d\mathbf{l} = (F_2 - F_1)b = |B_2 - B_1|qv_xb = -q\frac{d\Phi}{dt} \tag{8.1}$$

where $d\Phi/dt$ is the time rate of change of magnetic induction flux encircled by the loop. This result follows from the fact that $Bbv_x = Bb\,dx/dt$, and $B_2b\,dx$ is the flux added at the right-hand side of the loop in a time dt, while $B_1b\,dx$ is the amount of flux removed from the left-hand side in the time dt. The net rate of change of flux through the loop is thus $(B_2 - B_1)v_xb = -d\Phi/dt$. The negative sign comes from the fact that we have chosen the direction of B to be negative (pointing in, or away from the observer). As the loop moves to the right, the magnitude of the negative flux increases.

When we divide Eq. (8.1) by q, we find the emf,

$$\frac{W}{q} = \mathcal{E} = -\frac{d\Phi}{dt} \tag{8.2}$$

This emf will drive charges around the loop just as will a chemical cell. Thus when the resistance around the loop is R, the current is given by $\mathcal{E} = iR$. When a loop is moving in a region where the magnetic induction varies with position, there arises an emf that can be expressed in terms of the time rate of change of magnetic flux contained within the loop. We discuss the sign of the emf in relation to the flux change in Sec. 8.4.

8.3 Faraday's Law of Induction

The rule for the generation of an emf just found by an argument based on the magnetic force on moving charges can be made much more general. Figure 8.2 shows a coil (1) with galvanometer in

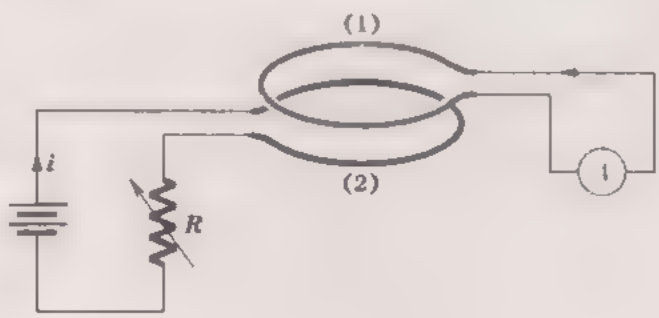


***Fig. 8.2** Induced emf via Faraday's law of induction.*

series, placed above another coil (2). Coil 2 is connected to a battery that causes a current i and therefore produces a magnetic induction field in the region of the two coils. When we lower coil 1 toward coil 2, we observe a current in (1) in a direction as shown. The geometry is a little different from the case previously studied,

but we could easily show that this induced current flow in (1) is exactly according to Eq. (8.2). The motion of coil 1 moves it into a region of greater B, resulting in an emf that could be calculated either on the basis of $d\Phi/dt$ within the coil or, more laboriously, by means of the magnetic force on charges in coil 1. If we move coil 1 away from rather than toward coil 2, the induced current in (1) is reversed as expected.

It is interesting to ask what happens if, instead of lowering coil 1 toward coil 2, we raise coil 2 toward coil 1. On the basis of our derivations of the induced emf, we can say very little, since the force on a charge at rest in a moving field is now involved a situation we have not investigated. There is, however, a very strong argument for believing that the resulting induced current in coil 1 will be the same, whether (1) is moved toward (2) or vice versa. If there *were* a difference between these two experiments, an absolute reference frame for motion could be established. In fact, the experimental result is that it makes no difference which coil is moved. *Only the relative motion is significant.* This is, incidentally, one of the principal postulates of Einstein's theory of special relativity, which says that there is no *absolute* frame of motion. All the laws of electricity and magnetism are *relativistically invariant;* that is, they give results that are always consistent with relativity. This is not true in the simple Newtonian formulation of the laws of mechanics.

A third kind of experiment can be done with the two coils. This time both coils are held at rest (or, more generally, held fixed with respect to each other) and the flux in coil 1 is changed by varying the current in coil 2. Here again, we cannot predict from prior knowledge whether or not the result for this kind of $d\Phi/dt$ is identical with that due to motion. The experimental fact is that the result is indeed identical.

The generalized statement, that

$$\mathcal{E} = -\frac{d\Phi}{dt} \tag{8.3}$$

is true around any closed path, regardless of the cause of the time rate of change of the magnetic flux threading the path, is the experimental fact known as Faraday's law of induction. It is consistent with the law of magnetic force on moving charges as discussed in the last section, but in its general form must be considered as an

independent law. A further aspect of Faraday's law is that it applies in any region of space, regardless of whether a conductor is present. This fact is essential in the propagation of electromagnetic waves.

Faraday's experiments on induced emf and the laws of induction were published between 1831 and 1854. In the following paragraph,[1] published in 1832, Faraday described one of the numerous experiments through which he developed his ideas of induced emf:

> Two hundred and three feet of copper wire in one length were passed round a large block of wood; other two hundred and three feet of similar wire were interposed as a spiral between the turns of the first coil, and metallic contact everywhere prevented by twine. One of these helices was connected with a galvanometer, and the other with a battery of one hundred pairs of plates of four inches square, with double coppers, and well charged. When the contact was made, there was a sudden and very slight effect at the galvanometer, and there was also a similar slight effect when the contact with the battery was broken. But whilst the voltaic current was continuing to pass through the one helix, no galvanometrical appearances nor any effect like induction upon the other helix could be perceived, although the active power of the battery was proved to be great, by its heating the whole of its own helix, and by the brilliancy of the discharge when made through charcoal.

He summarizes the results of a whole group of such experiments involving both changing currents and moving magnets, in the following short paragraph,[2] which, allowing for a considerable change in nomenclature since his time, we see is referring to the existence of induced emf:

> All these results show that the power of inducing electric currents is circumferentially excited by a magnetic resultant or axis of power, just as circumferential magnetism is dependent upon and is exhibited by an electric current.

8.4 Lenz's Principle

We now come to a discussion of the relationship of the direction of the induced emf with respect to the sign of the rate of change

[1] Michael Faraday, Experimental Researches in Electricity, *Phil. Trans. Roy. Soc. London,* (A) **122:** pp. 127, 155 (1832).
[2] *Ibid.*

of flux. In Sec 8.2, the discussion of motional emf showed that a negative sign appears in the equation when the convention of positive current and emf for counterclockwise rotation is used, and when positive flux is directed outward toward the viewer. This is true for all cases where Faraday's law is involved. This negative relationship between induced emf and $d\Phi/dt$ follows directly from the principle of energy conservation, which in this application is often called Lenz's principle. We use the example of Sec. 8.2 to determine the consequences of this rule. There the motion of the loop increases the magnitude of the flux Φ, which is itself negative since it points inward. Thus $d\Phi/dt$ is negative, and the emf is consequently positive or counterclockwise. Any current produced in the loop by the magnetic forces is thus positive and so produces an additional flux Φ', which is positive or outward. Thus the induced flux opposes the time rate of change of the externally produced flux. We can always calculate the direction of the induced emf on this basis.

We can show that this behavior is required on the basis of energy conservation by considering the alternative possibility. Suppose we have a conducting loop in a space that originally contains no magnetic induction field. Let there be a small field imposed from outside (or suppose a small fluctuation in the thermal motions of electrons in the loop produces a very small current for an instant, thereby generating a very small magnetic induction field in the loop). If the reverse of Lenz's principle were true, this change from zero field to a small field would generate an emf and a consequent current that would still further increase the magnetic flux in the same direction. The increasing rate of change of flux would produce still more current causing more increase, thereby causing an almost explosive generation of very high current and very high field. The end would presumably come when the current in the wire becomes so great that the energy dissipation causes the wire to melt. Such a series of events clearly contradicts the principle of energy conservation, so we have in effect shown that Lenz's principle must hold. Therefore the induced emf must always act in the stated "contrary" fashion.

8.5 *Examples*

a Conservation of energy in an elementary generator We study the energy aspect of induced emf by considering a very elementary electric generator involving Faraday induction. Figure 8.3 shows

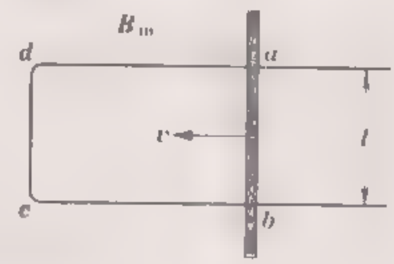


Fig. 8.3 *Sliding wire on a stationary loop in a field B.*

a sliding wire that can be moved on a stationary loop in a uniform field B. We pull the rod, as shown, with a uniform velocity v. This motion will result in a change in the magnetic flux through the loop and the generation of an emf,

$$\frac{d\Phi}{dt} = vlB = -\mathcal{E}$$

Here $d\Phi/dt$ is positive since Φ is inward and is decreased in magnitude as the area of the loop is reduced. This means the emf is negative, or clockwise. (We could have calculated this emf by means of the method of Sec. 8.2.) As a result of this emf, a current flows,

$$i = \frac{\mathcal{E}}{R} = \frac{Blv}{R}$$

where R is the resistance around the loop.

Electric energy is being dissipated in the resistance at a rate

$$P = \mathcal{E}i = \frac{B^2l^2v^2}{R} \qquad \text{watts}$$

The source of this energy is the mechanical work being done in moving the rod. The force on the rod carrying a current i in a field B is $F = Bil$, so that the rate at which mechanical work is being done is

$$\frac{dW}{dt} = Fv = Blvi = \frac{B^2l^2v^2}{R} \qquad \text{watts}$$

We have shown the equality of mechanical work rate and electric energy dissipation rate.

b Mechanical work done in moving a coil in a magnetic induction field We determine the mechanical work done when we move a coil carrying a constant current i in a magnetic field (Fig. 8.4).

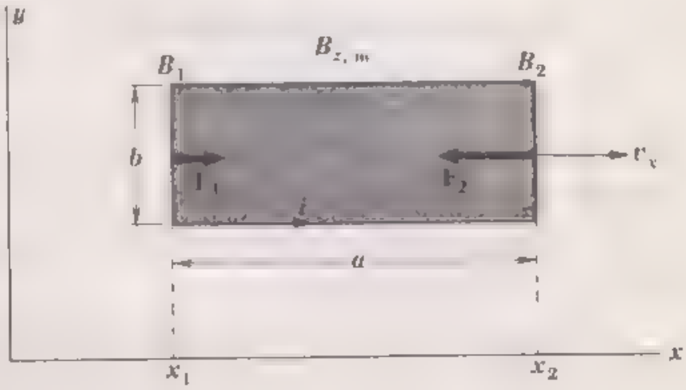


Fig. 8.4 Work done in moving a loop in a magnetic induction field.

We use this result later in the calculation of mutual inductance (Sec. 8.6). Assume the current is driven by an external current source in a positive (counterclockwise) direction. As a result of this current, there are forces F_1 and F_2 acting on the sides as shown, where

$$F_1 = B_1 ib \qquad \text{and} \qquad F_2 = B_2 ib$$

If B is perpendicular to the xy plane and varies only with x, the forces on the top and bottom sections of the loop are equal and opposite and therefore cancel. In any case, we shall move the loop in the x direction, so these latter forces in the y direction do not contribute to the work done. To be definite, we assume that the magnitude of B increases with x, so $B_2 > B_1$. The net magnetic force on the loop is $-F = -(F_2 - F_1)$, to the left. The negative sign is used because the force is in the negative x direction.

The external work done in moving the loop to the right a distance dx is

$$dW = F\,dx = (F_2 - F_1)\,dx = ib|B_2 - B_1|\,dx \tag{8.4}$$

We wish to connect this external mechanical work done to the net change in the externally produced magnetic induction flux through the loop. The change in flux when the loop is moved a distance dx is

$$b|B_2 - B_1|dx = -d\Phi$$

The negative sign is used because the displacement of the loop in the positive x direction increases the magnitude of Φ through the loop, and in this case Φ is pointed away from the observer and is therefore a negative quantity. Comparison of this equation with (8.4) gives

$$dW = -i\,d\Phi \tag{8.5}$$

For N turns in the coil, the mechanical work done is

$$dW = -Ni\,d\Phi \tag{8.6}$$

Here the usual sign convention applies: i counterclockwise is positive and Φ is positive when it is outward, toward the observer. In the example chosen, since i is positive and $d\Phi$ is negative, the external work done is positive.

The calculation above is confined to the *mechanical* work done in moving a current loop in a magnetic field. There is of necessity some electric work done also. The source of emf producing the current i in the coil must do extra work beyond the i^2R term in order to compensate for the induced emf in the loop caused by the changing flux in the loop. In addition, the motion of the current loop produces a changing flux in the external circuit acting as the source of external magnetic field and requires work by this source to compensate for the induced emf in the external circuit. This electric work need not concern us at present.

c An elementary motor We use the result of the last problem in an examination of a simple motor that can convert electric into mechanical energy. Figure 8.5 shows the device we shall discuss

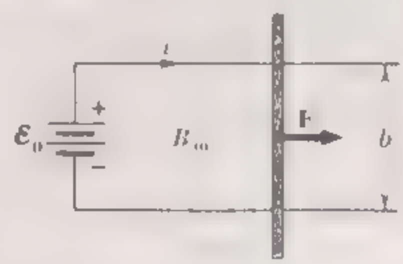


***Fig. 8.5** An elementary motor.*

The chemical cell of emf $\mathcal{E}_0$ causes a current i to flow through the loop whose total resistance is R. The magnetic force on the movable rod is

$$F = ibB = \frac{\mathcal{E}_0}{R}bB$$

The rate at which work is done by the magnetic force F, when the rod moves at a velocity v, is

$$\frac{dW}{dt} = Fv = i\frac{d\Phi}{dt} \tag{8.7}$$

where we have used Eq. (8.6) to relate work to flux change. The negative sign is omitted because, while $d\Phi/dt$ is positive, i is clockwise or negative. The induced emf caused by $d\Phi/dt$ must be included in the circuit equation, which may be written as

$$\mathcal{E}_0 - \frac{d\Phi}{dt} = iR \tag{8.8}$$

If this equation is multiplied by i, we get

$$\mathcal{E}_0 i = i\frac{d\Phi}{dt} + i^2R = Fv + i^2R \tag{8.9}$$

Thus th battery supplies energy at a rate $\mathcal{E}_0 i$, producing mechanical work at the rate Fv and dissipating energy in the resistance of the circuit at the rate i^2R. The motor is discussed in more detail in Fxample 8.8d and Sec. 10.13.

8.6 Mutual Inductance

We now undertake a discussion of the interaction between two circuits by virtue of magnetic flux linkage. As a result of flux linkage, combined with the Faraday induction law, changes in one

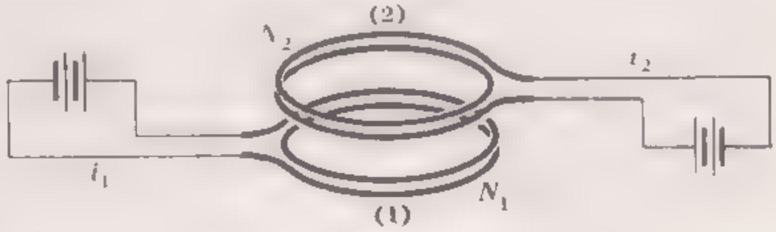


Fig. 8.6 Mutual inductance between two circuits.

circuit are felt by the other. We formulate this interaction in terms of the *mutual inductance* between two circuits. We begin the study by examining the energy involved in bringing two magnetic circuits together. Figure 8.6 shows two coils, carrying constant currents i_1 and i_2 and having N_1 and N_2 turns, respectively. We start with the two coils widely separated and ask for the amount

of external work that must be done to bring coil 1 up to the vicinity of coil 2. Referring back to Example 8.5*b*, we can write

$$dW = -N_1 i_1 \, d\Phi_{12} \tag{8.6}$$

where $d\Phi_{12}$ is the additional flux that links coil 1 by virtue of the current in coil 2, when the coils are brought closer together. The total mechanical work done in bringing the coils together from far apart is then

$$W = \int_{\infty}^{\text{final position}} dW = -N_1 i_1 \int_{\infty}^{\text{final position}} d\Phi_{12} = -N_1 i_1 \Phi_{12} \tag{8.10}$$

We could equally well have arrived at the final state by holding coil 1 fixed and moving coil 2 up from infinity. This would give

$$W = -N_2 i_2 \Phi_{21} \tag{8.11}$$

Since the final states are equivalent regardless of the way in which they are achieved, these two energies must be equal, so we may write

$$N_2 i_2 \Phi_{21} = N_1 i_1 \Phi_{12}$$

or

$$\frac{N_1 \Phi_{12}}{i_2} = \frac{N_2 \Phi_{21}}{i_1} \tag{8.12}$$

These ratios are called the mutual inductance M between the two coils. $N_1\Phi_{12}$ is the *flux linkage* in coil 1 due to current in coil 2, that is, the flux in (1) due to current in (2) times the number of turns in (1). The mutual inductance is simply *the flux linkage in one coil due to unit current in the other coil.* We have proved that $M_{12} = M_{21}$, regardless of the geometries of the individual coils. Here we are defining

$$M_{12} = \frac{N_1 \Phi_{12}}{i_2} \qquad \text{and} \qquad M_{21} = \frac{N_2 \Phi_{21}}{i_1} \tag{8.13}$$

but we have now seen that they have the same value.

We now use the mutual inductance between two coils to evaluate the effect of changing the current in one coil on the behavior of the current in the second coil. The two coils of Fig. 8.6

are held fixed in space and the current is changed in, say, coil 2. There results an emf induced in coil 1, given by

$$\mathcal{E}_1 = -N_1 \frac{d\Phi_{12}}{dt}$$

Since $\Phi_{12} = M i_2/N_1$, from Eq. (8.13) we can find the change in Φ_{12} when i_2 is changed by differentiating to get

$$d\Phi_{12} = \frac{M}{N_1}\, di_2$$

Substitution above gives

$$\mathcal{E}_1 = -M \frac{di_2}{dt} \qquad \text{volts} \tag{8.14}$$

Similarly, if i_1 is varied, the emf induced in coil 2 is

$$\mathcal{E}_2 = -M \frac{di_1}{dt} \qquad \text{volts} \tag{8.15}$$

The unit of mutual inductance in both the mks and the practical system of units is the *henry*, named in honor of the American scientist Joseph Henry, who developed the idea of inductance almost simultaneously with Faraday. The mutual inductance is unity if a change of 1 amp/sec in one coil induces an emf of 1 volt in the other coil. Thus one henry is one volt per ampere per second. We give an example of a simple calculation of the effect of mutual inductance in Sec. 8.8.

8.7 Self-inductance

Faraday induction is of equal importance in a single circuit, as can be seen by examining the effect in a single coil such as shown in Fig. 8.7. If the current in this coil is changed (say by varying

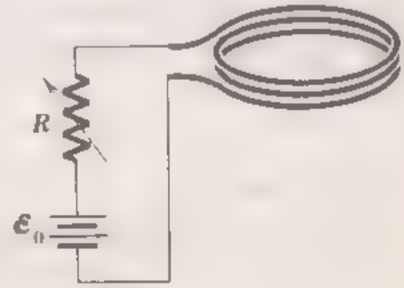


Fig. 8.7 Self-inductance in a coil.

the resistance R), the flux threading the coil changes, and there is a consequent emf induced in it. By analogy with the definition of mutual inductance, the self-inductance L may be defined as

$$L = \frac{N_1\Phi_{11}}{i_1} \tag{8.16}$$

Here Φ_{11} is the flux in the coil due to its own current, so the self-inductance is *the flux linkage in a circuit per unit current in that circuit*. As above, we find for the emf induced by a changing current,

$$\mathcal{E}_1 = -\frac{N_1 d\Phi_{11}}{dt}$$

From Eq. (8.16) we have

$$\Phi_{11} = \frac{L_1 i_1}{N_1}$$

so

$$d\Phi_{11} = \frac{L_1}{N_1}\,di_1$$

Substitution in the equation for $\mathcal{E}_1$ then gives

$$\mathcal{E}_1 = -L\frac{di_1}{dt} \qquad \text{volts} \tag{8.17}$$

Any closed circuit must have a self-inductance, even if no coil is included, since a complete circuit involves at least one loop that contains some flux when current flows. However, if the area of the loop is made very small, the flux linkage will also be very small, and the self-inductance is much reduced. The unit of self-inductance is also the henry.

Before continuing our discussion, we may note that the effect of inductance in a circuit is like inertia in a mechanical system. Thus any change in a steady current is accompanied by an induced emf that tends to counteract this change. This is analogous to the force exerted by a moving mass when we try to modify its velocity.

8.8 Examples

a Effect of mutual inductance in coupled circuits We examine the quantitative behavior of two circuits coupled by mutual in-

ductance. As shown in Fig. 8.8, the current in circuit 2 can be changed by varying R_2. We calculate the effect on the current in circuit 1 of a given value of di_2/dt in circuit 2. Using Eq. (8.15),

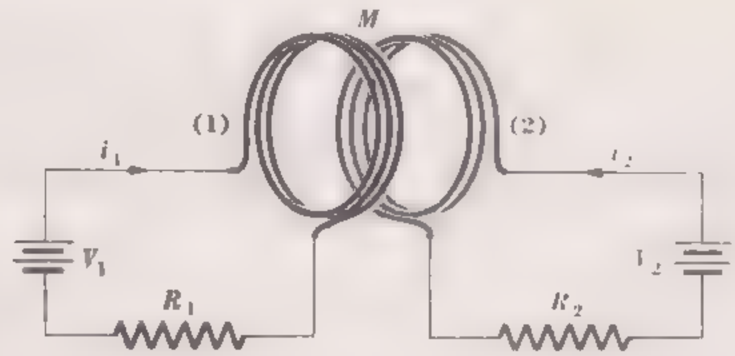


***Fig. 8.8** Two circuits coupled by mutual inductance M.*

we write directly for the induced emf in circuit 1 due to the mutual inductance,

$$\mathcal{E}_1 = -M \frac{di_2}{dt} \qquad \text{volts}$$

The negative sign tells us that if the current in the two circuits is going around the coils in the same sense, $\mathcal{E}_1$ will be opposed to the source V_1 when di_2/dt is positive. When the polarity of V_1 is reversed or when one of the coils is rotated 180°, $\mathcal{E}_1$ and V_1 act in the same direction when di_2/dt is positive.

One additional effect to be considered is that of L_1, the self-inductance in circuit 1. This produces another emf in circuit 1, according to Eq. (8.17),

$$\mathcal{E}_1' = -L \frac{di_1}{dt} \qquad \text{volts}$$

which acts in opposition to di_1/dt. The equation for circuit 1 becomes

$$V_1 \pm M \frac{di_2}{dt} - L_1 \frac{di_1}{dt} = i_1 R \tag{8.18}$$

If the value of di_2/dt and the sense in which M acts on circuit 1 are given, this differential equation can be solved by methods similar to those discussed in Sec. 10.7. If di_2/dt is not constant, the problem is more complicated. It has not been necessary to include the self-inductance L_2, since we have postulated a certain value of

di_2/dt. The effect of L_2 is to produce an emf in circuit 2 that must be compensated for when we vary R_2 to cause the rate of change of i_2.

b Calculation of self-inductance In a few cases it is easy to calculate the self-inductance of particular geometric arrangements. We make an illustrative calculation for the case of a uniform long coil, neglecting end effects that in practice would reduce the inductance somewhat, owing to the lower flux at the ends of the coil. We expect the number of turns N to enter as N^2, once as the N in $N\Phi$ and once again since Φ itself is proportional to N. We have already calculated in Examples 6.3*c* and 6.6*b* the field B inside a solenoid of length l and N turns. We found $B = \mu_0 Ni/l$. Substitution in Eq. (8.16) gives for the self-inductance

$$L = \frac{\mu_0 N^2 A}{l} \qquad \text{henrys} \tag{8.19}$$

If this is a toroidal solenoid, no flux leaks out and the result is accurate.

c Calculation of emf of a generator A more practical generator for converting mechanical work into electric energy than the one discussed in Example 8.5*a* is shown in Fig. 8.9. A coil of N turns

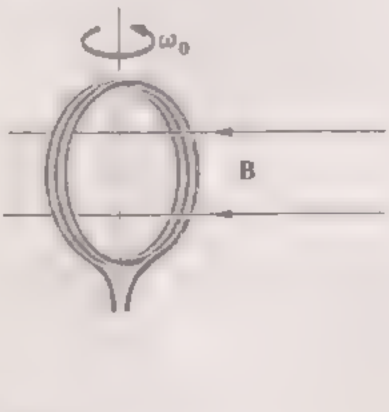


***Fig. 8.9** A simple generator: a rotating coil in an external magnetic induction field.*

and area A is rotated in a uniform magnetic induction field B, at an angular velocity ω_0 radians/sec. We calculate the emf generated using the Faraday induction law. The flux linkage threading the coil at an angle θ between the plane of the coil and the direction of B is given by

$$N\Phi = NAB\cos\theta = NAB\omega_0 t$$

where $\omega_0 t$ is the value of θ for a particular time t. Then

$$\mathcal{E} = -N\frac{d\Phi}{dt} = \omega_0 NAB \sin \omega_0 t \tag{8.20}$$

Thus this generator gives a sinusoidal emf of amplitude $\omega_0 NAB$. The source of this emf is the mechanical work done in rotating the coil. This is zero (in a frictionless system) if no current flows and increases linearly with increasing current, as found in Example 8.5*a*. A discussion of several types of generators is given in Sec. 10.12.

d Calculation of back emf in a motor A more practical motor than the one discussed in Example 8.5*c* is the one sketched in Fig. 8.10. Here the torque acting on a current-carrying loop in a

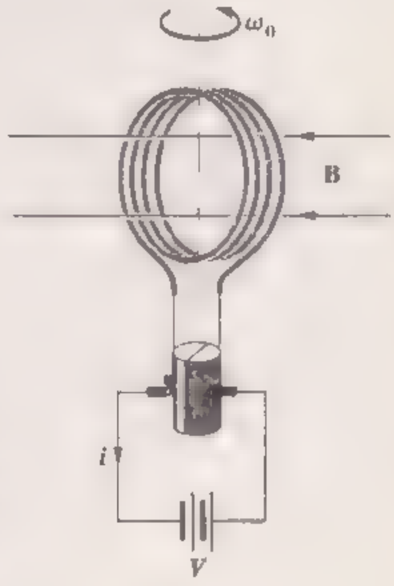


***Fig. 8.10** A simple motor: a current-carrying coil with commutator, in external magnetic induction field.*

magnetic induction field is used to cause rotation of the coil. A *commutator* is used to reverse the current at the appropriate phase of rotation so that the torque always acts in the same direction. Let us suppose that the angular velocity of the coil is ω_0, with some particular external mechanical load. Just as in the case of the generator, the rotation of the motor coil in the magnetic field results in an induced emf. Since this emf acts to oppose the flow of current in the motor coil, it is called a back emf and is given by

$$\mathcal{E} = -N\frac{d\Phi}{dt} \tag{8.21}$$

where N is the number of turns in the coil. If the magnetic induction field strength is B_0 and the coil area is A,

$$\Phi = AB_0 \sin \omega_0 t$$

and

$$\frac{d\Phi}{dt} = AB_0\omega_0 \cos \omega_0 t$$

The commutator changes the $d\Phi/dt$ curve from $\cos \omega_0 t$ to one like that in Fig. 8.11. The back emf thus varies in a similar fashion,

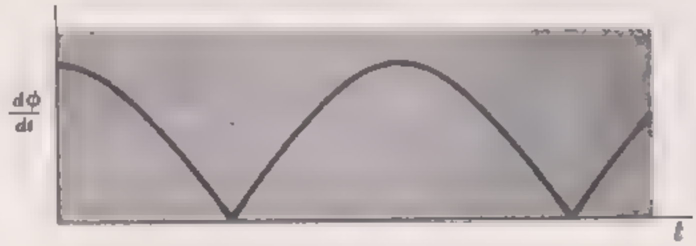


Fig. 8.11 Back emf of motor as seen through commutator.

according to Eq. (8.21). If we call this back emf which varies periodically $\mathcal{E}(t)$, the current in the coil is given by the expression

$$V - \mathcal{E}(t) = iR$$

and we see that i will also vary with time, in a way indicated in

Fig. 8.12 Variation of current in motor due to back emf.

Fig. 8.12. In the figure, i_0 is the current that would flow if the loop were held stationary. Here we have neglected the further modi-

fications that would occur if the self-inductance of the coil were important at the frequency ω_0. The average current is given by

$$V - \overline{\mathcal{E}(t)} = iR$$

where $\overline{\mathcal{E}(t)}$ is the average value of the induced emf.

In practice, rather than using a single coil, the rotating unit of a motor, called an *armature*, is made up of a number of coils, which are oriented at different angles around the axis of rotation. The commutator connects the external source of power to each coil consecutively in such a way that maximum torque is obtained from each coil in turn.

The back emf plays an important role in the operation of a motor Let us explain this by considering a motor armature in which the coils have very little resistance. If a power source is connected to the coils, a large current flows while the coils are at rest, limited only by the resistance of the coils. However, as soon as the coils begin to turn in the static magnetic field produced by stationary coils in the motor, a back emf is induced that opposes the externally applied voltage causing the current in the rotating coils. If the motor is running without a mechanical load, the back emf reduces the current to a small value, just large enough to supply the energy used up by friction and electrical resistance losses. A heavy mechanical load tends to slow down the motor, causing a reduction in the back emf that allows enough electric power to be used to account for the greater mechanical load. On this basis, one can understand the large electric power drain when starting a motor. A discussion of a number of types of motors is given in Sec. 10.13.

8.9 *Calculation of Mutual Inductance*

There is one situation in which it is easy to calculate the mutual inductance between two circuits in terms of the self-inductance of each circuit alone. This is the case of maximum flux linkage, when all the flux through one circuit links the other. One physical arrangement that would achieve this would be to place two coils of equal area (but not necessarily of equal numbers of turns) next to each other. Another arrangement giving maximum flux linkage would be two coils wound one on top of the other on a toroidal

solenoid. Calling the coils 1 and 2, we can write for the mutual inductance

$$M = \frac{N_1\Phi_{12}}{i_2} = \frac{N_2\Phi_{21}}{i_1} \tag{8.12}$$

Similarly, the self-inductance of the coils is given by

$$L_1 = \frac{N_1\Phi_1}{i_1} \quad \text{and} \quad L_2 = \frac{N_2\Phi_2}{i_2} \tag{8.16}$$

Since all the flux of coil 2 links coil 1,

$$\frac{\Phi_{12}}{i_2} = \frac{\Phi_2}{i_2}$$

that is, the flux Φ_{12} in coil 1 due to the current i_2 is the same as the flux Φ_2 in coil 2 due to the same current i_2. Similarly,

$$\frac{\Phi_{21}}{i_1} = \frac{\Phi_1}{i_1}$$

If we replace Φ_{12}/i_2 and Φ_{21}/i_1 in the equations for M above and multiply M_{21} by M_{12}, we get

$$M_{12}M_{21} = M^2 = \frac{N_1N_2\Phi_1\Phi_2}{i_1i_2} = L_1L_2$$

or

$$M_{max} = \sqrt{L_1L_2} \tag{8.22}$$

This represents the maximum possible value of M in terms of L_1 and L_2. In general, the mutual inductance is given by

$$M = k\sqrt{L_1L_2} \tag{8.23}$$

where k is a number between 0 and 1, which depends on the geometry of the coils and their relative positions.

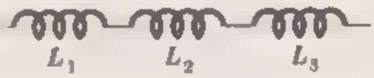
8.10 Combinations of Inductances

Several inductances in series, if so arranged that there are no interactions through mutual inductance, are easily shown to behave as a single inductance equal to their sum. Thus the induced voltage across the three isolated inductances shown in Fig. 8.13 will be

$$-L_1\frac{di}{dt} - L_2\frac{di}{dt} - L_3\frac{di}{dt} = -L\frac{di}{dt}$$

where $L = L_1 + L_2 + L_3$. However, the situation is modified

Fig. 8.13 Three self-inductances in series, no interactions between them. Equivalent value, $L = L_1 + L_2 + L_3$.



if the flux from one inductance links another, giving rise to mutual-inductance terms. Figure 8.14 shows two coils that interact in this fashion. The voltage across this pair of coils will be given by

$$\mathcal{E} = -L_1\frac{di}{dt} - L_2\frac{di}{dt} \pm 2M\frac{di}{dt} \tag{8.24}$$

The mutual inductance occurs twice, once for the voltage induced in the second coil by current changes in the first and once for the

Fig. 8.14 Interacting self-inductances in series. Equivalent value, $L = L_1 + L_2 \pm 2M$.

converse. The $\pm$ allows for the possibility that the mutual interaction adds to the total flux or that it gives flux in the opposite direction from that due to the self-induction. The sign of M and its magnitude depend on the geometric arrangement of the coils. From Eq. (8.24) we find for the equivalent self-inductance of the pair of coils in series:

$$L = L_1 + L_2 \pm 2M \tag{8.25}$$

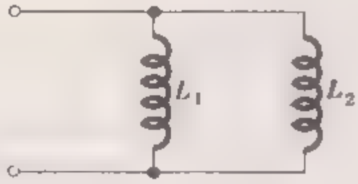


***Fig. 8.15** Two inductances in parallel.*

For the case of noninteracting inductances in parallel as shown in Fig. 8.15, we may write

$$\mathcal{E}_1 = -L_1 \frac{di_1}{dt}$$

$$\mathcal{E}_2 = -L_2 \frac{di_2}{dt}$$

But in the parallel arrangement, $\mathcal{E}_1 = \mathcal{E}_2$, and also

$$\frac{di}{dt} = \frac{di_1}{dt} + \frac{di_2}{dt} = -\left(\frac{1}{L_1} + \frac{1}{L_2}\right)\mathcal{E}$$

The equivalent inductance L is given by $\mathcal{E} = -L\,di/dt$, so we have

$$\frac{1}{L} = \frac{1}{L_1} + \frac{1}{L_2} \qquad \text{or} \qquad L = \frac{L_1L_2}{L_1 + L_2} \tag{8.26}$$

Note that these rules for combining noninteracting inductances turn out to be of the same form as the rules for series and parallel combinations of resistances.

8.11 Stored Magnetic Energy

The induced emf in a circuit resulting from changing magnetic flux can be shown to lead to the expenditure of energy in order to set up a magnetic field in a region of space. Since this expended

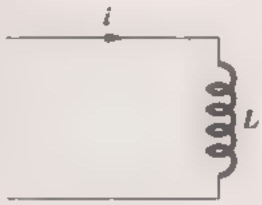


***Fig. 8.16** Buildup of current in an inductance.*

energy can be recovered and used, for instance, to generate heat in a resistance, we are led naturally to the idea of energy stored in a magnetic field. The situation is analogous to the storage of

energy in an electric field, as discussed in Chap. 4. We begin the discussion by asking how much energy must be expended to build up a current i in an inductance L_1 as shown in Fig. 8.16. The induced emf across the inductance is

$$\mathcal{E} = -L\frac{di}{dt}$$

The amount of work required to move a charge dq against this emf is

$$dW = -\mathcal{E}dq = L\frac{di}{dt}dq = L\frac{dq}{dt}di = Li\,di \tag{8.27}$$

Here we have changed variables and used $i = dq/dt$. The total work to build up the current from 0 to i is obtained by integrating Eq. (8.27):

$$U = \int dW = L\int_0^i i\,di = \tfrac{1}{2}Li^2 \tag{8.28}$$

We have written this as the stored energy U since, as we shall show, the energy expended can be recovered.

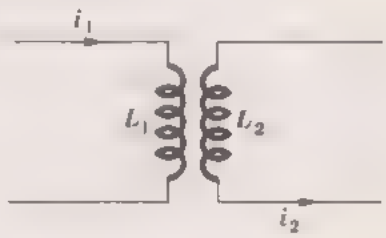


Fig. 8.17 Energy stored in two interacting circuits.

If two interacting circuits are involved as in Fig. 8.17, the energy required to build up the currents i_1 and i_2 in the two circuits is obtained similarly. Thus,

$$\begin{aligned} dW &= -\mathcal{E}_1\,dq_1 - \mathcal{E}_2\,dq_2 \\ &= \left(L_1\frac{di_1}{dt} \pm M\frac{di_2}{dt}\right)dq_1 + \left(L_2\frac{di_2}{dt} \pm M\frac{di_1}{dt}\right)dq_2 \\ &= L_1i_1\,di_1 \pm Mi_1\,di_2 + L_2i_2\,di_2 \pm Mi_2\,di_1 \end{aligned}$$

Since $Mi_1\,di_2 + Mi_2\,di_1$ is a perfect differential, these two terms can be written as $d(Mi_1i_2)$. Integration of the entire expression then yields

$$U = \tfrac{1}{2}L_1i_1^2 + \tfrac{1}{2}L_2i_2^2 \pm Mi_1i_2 \tag{8.29}$$

We can show that this energy is stored rather than dissipated by arranging that the stored energy be used to heat up the resistance. Suppose, then, that while the current i is flowing steadily we instantaneously remove from the circuit the battery producing the current. The circuit becomes that shown in Fig. 8.18. This

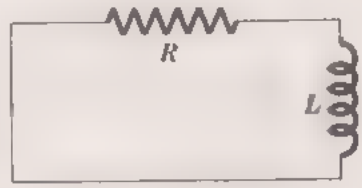


***Fig. 8.18** Dissipation of energy stored in the magnetic field of an inductance.*

change can be made essentially instantaneously by simply shorting out the battery. In the absence of the inductance, the current in the circuit would fall to zero immediately the battery was removed by shorting. However, with the inductance present, the circuit equation becomes

$$-L\frac{di}{dt} = iR \tag{8.30}$$

The inductance acts so as to try to maintain the current constant. Since in its absence the current would decrease, di/dt is negative, and we see from Eq. (8.30) that the induced emf acts in the same direction as did the battery. The inductance thus supplies the energy to force the current through the resistance. The instantaneous rate of energy dissipation will be

$$-Li\frac{di}{dt} = i^2R \tag{8.31}$$

obtained from Eq. (8.30) by multiplying by the instantaneous current i. Since in this situation i and di/dt have opposite signs, the left-hand term is positive. The total energy supplied to the resistance is the integral of this expression over the time while both i and di/dt are not zero. We have already found this to be $\frac{1}{2}Li^2$. This is equal to $\int i^2R\,dt$, the total energy dissipated in the resistance after the battery is removed from the circuit.

We have now demonstrated that an amount of energy $\frac{1}{2}Li^2$ is necessary to build up a current i in an inductance and that this energy is stored. This energy is associated with the existence of

current in the inductance. A more useful point of view, however, is that the energy is stored in the magnetic field existing in space because of the current. This is closely analogous to the situation in the case of energy stored in an electrostatic field, where we found $U/\text{vol} = \frac{1}{2}\epsilon_0 E^2$ in a vacuum. A rather similar calculation gives the energy storage in a magnetic field. We choose a toroidal solenoid because the field exists only inside the toroid and also is essentially uniform within it, making the calculation easy. We have already found that the inductance of such a solenoid is $L = \mu_0 N^2 A/l$ henrys. The total energy stored in the field is then $U = \frac{1}{2}Li^2 = \frac{1}{2}\mu_0 N^2 A i^2/l$ joules. The energy per unit volume is the energy divided by the volume inside the solenoid, Al. This gives $U/\text{vol} = \frac{1}{2}\mu_0 N^2 i^2/l^2$. But since B inside is $B = \mu_0 Ni/l$, we find

$$\frac{U}{\text{vol}} = \frac{1}{2}\frac{B^2}{\mu_0} \qquad \text{joules/m}^3 \tag{8.32}$$

for the case of a magnetic field B in a vacuum.

8.12 Further Discussion of Faraday's Law of Induction

There are some aspects of Faraday's law of induction that have not yet been discussed. So far we have been interested in the emf induced around a conductor placed in a region where B is changing with time. The emf produces a force on charged particles in the conductor and so gives rise to an electric field, which can be related to the induced emf $\mathcal{E}$ by the equation

$$\mathcal{E} = \oint \mathbf{E}\cdot d\mathbf{l} = \oint E_{tan}\, dl \tag{8.33}$$

When we take a circular conductor (Fig. 8.19) whose plane is

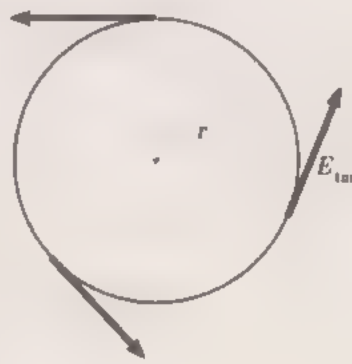


Fig. 8.19 Tangential electric field, induced along a circular wire in a region of changing B field.

perpendicular to the lines of a uniform B field, this tangential electric field is uniform around the circuit and is given by

$$E_{tan} = \frac{\mathcal{E}}{2\pi r} \tag{8.34}$$

where r is the radius of the circular path. The work done on a charge q during one revolution of the charge is

$$W = \oint qE_{tan}\,dl = qE_{tan}\,2\pi r = q\mathcal{E} \qquad \text{coulomb} \times \text{volts} \tag{8.35}$$

Notice that the direction of this electric field is not defined. The direction of E_{tan} depends entirely on the path chosen, so it is really quite different from an electrostatic E field. Another difference is that for this induced field, $\oint E \cdot dl \neq 0$.

Another important property of the induced emf is that it exists whether or not a conducting path is present. This fact is well exemplified in a kind of electron accelerator called the *betatron*. In this machine, electrons in a vacuum travel in circular paths in a magnetic field. The strength of the magnetic field changes with time, and as a result of the changing magnetic flux in the area enclosed by the electron orbit, an emf is induced that accelerates the electrons.

8.13 Applications of Inductance

In Chap. 10 we discuss inductances as circuit elements in transient and a-c circuits. Here we discuss two applications of induced emfs to practical problems.

The ballistic galvanometer Perhaps the most popular method for measuring the strength of a constant magnetic field is the use of a *flip coil.* A small coil of known geometry is placed in the unknown field so as to be threaded by the maximum flux. Then the coil is suddenly removed from the field, or rotated so as to reduce the flux through it to zero. If the rate of change of flux that occurs could be integrated from beginning to end, the flux and hence B could be obtained. Thus,

$$\int \frac{d\Phi}{dt}\,dt = \Phi = BA$$

where Φ is the value of the original flux through the coil. We now show that under suitable conditions, this change in flux can be related to a galvanometer deflection caused by the induced emf in the flip coil, $\mathcal{E} = -N\,d\Phi/dt$.

The method involves the use of the *ballistic galvanometer*, a galvanometer built with its natural period of mechanical oscillation reasonably long. The galvanometer is connected across the coil as shown in Fig. 8.20. When the magnetic flux in the coil is suddenly

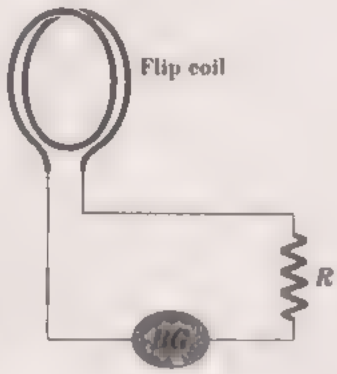


***Fig. 8.20** Flip coil used with ballistic galvanometer to measure B.*

changed, the galvanometer coil is given an angular impulse, the magnitude of which is proportional to the total change in flux. If the galvanometer characteristics are known, the absolute magnitude of the field can be determined.

The principle involved is as follows: The magnetic torque on a galvanometer coil, when its plane is parallel to the magnetic field, we have found to be

$$\tau = NiAB \tag{8.36}$$

where the quantities $NiAB$ refer to the galvanometer coil, not the flip coil. When, instead of a steady current i, we supply a short

***Fig. 8.21** Typical current pulse through ballistic galvanometer. Pulse must be finished before appreciable motion occurs in galvanometer.*

pulse of current such as shown in Fig. 8.21, the angular impulse given to the galvanometer coil is $\int \tau\,dt$, where the integration is

over an interval covering the entire time during which current flows. Using Eq. (8.36), we integrate this to get

$$\int \tau \, dt = NBA \int i \, dt = NBAQ \tag{8.37}$$

We have used the fact that $\int i \, dt$ is the total charge through the galvanometer. This is the special feature of the ballistic galvanometer. It receives an angular impulse that depends only on the total charge pulse flowing through it and not on the way the charge flow varies with time. However, since in Eq. (8.36) the more general expression is $\tau = NiAB \sin \theta$, where θ is the angle between the coil and field direction, the pulse must be completed before the coil has time to move from its initial position where $\sin \theta = 1$. The pulse must therefore be short compared with the natural period of the galvanometer coil. From mechanics we have that this angular impulse gives the coil an angular momentum $I\omega_0$, where I is its moment of inertia and ω_0 is the angular velocity just after the pulse has passed. Thus we have

$$\int \tau \, dt = I\omega_0 = NBAQ \tag{8.38}$$

The initial angular momentum $I\omega_0$ corresponds to an initial kinetic energy $\frac{1}{2}I\omega_0^2$. As the coil turns against the restoring torque of the galvanometer coil suspension, this kinetic energy is converted into potential energy. At the maximum deflection θ_{max}, conversion is complete and we may write

$$KE = \tfrac{1}{2}I\omega_0^2 \rightarrow PE = \tfrac{1}{2}k\theta^2{}_{max} \tag{8.39}$$

where k is the torque constant of the suspension. Solving for ω_0 in Eq. (8.38) and substituting in Eq. (8.39), we find

$$Q = \frac{\sqrt{Ik}}{NBA}\theta_{max} \tag{8.40}$$

This equation can be put in a more useful form by substituting the expression for the period of oscillation T of a torsional pendulum, $T = 2\pi(I/k)^{1/2}$, in Eq. (8.40):

$$Q = \frac{kT}{2\pi NAB}\theta_{max} = \frac{K'T}{2\pi}\theta_{max}$$

K' is the galvanometer constant (Sec. 7.7).

In terms of the figure of merit of the galvanometer [Eq. (7.31)], this becomes

$$Q = \frac{KT}{2\pi} d_{\max} \tag{8.41}$$

where $d_{\max}$ is the initial deflection of the galvanometer in millimeters on a scale 1 m away. If the galvanometer is appreciably damped, correction must be made for the effect of damping on the initial deflection.

We have now seen that knowledge of the galvanometer constant or figure of merit and of the period of oscillation of the galvanometer coil allows us to measure the total charge passing through the galvanometer in a time short compared with its period. We now show how this property of the ballistic galvanometer is used to measure the value of B. The current flowing in the flip coil connected across the galvanometer is given by

$$i = \frac{\text{emf}}{R} = -\frac{1}{R}\left(N'\frac{d\Phi}{dt} + L\frac{di}{dt}\right) \tag{8.42}$$

Here L represents the self-inductance in the circuit, and N' is the number of turns in the flip coil. Upon integration, we find the total flow of charge to be

$$Q = \int_0^\infty i\,dt = -\frac{N'}{R}(\Phi_2 - \Phi_1) - \int_0^0 L\,di = \frac{N'}{R}(\Phi_2 - \Phi_1) \tag{8.43}$$

where $(\Phi_2 - \Phi_1)$ is the change of the flux involved. If we pull the coil entirely out of the field, Φ_2 becomes zero and we have

$$Q = -\frac{N'}{R}\Phi_1 = \frac{N'}{R}BA' \qquad \text{or} \qquad B = \frac{QR}{N'A'} \tag{8.44}$$

where A' is the area of the search coil. Here it is assumed that the coil was originally oriented with its plane perpendicular to the magnetic field.

Eddy currents and skin depth Whenever a conductor is placed in a region of varying magnetic field, Faraday induction sets up currents. These currents result in i^2R losses in the metal, causing

heating and energy loss that may be undesirable. It is for this reason that in designing transformers the ferromagnetic metal cores usually used are made up of thin insulated slabs of metal (laminated) rather than of solid metal. To show how this cuts down on losses, we make a simple approximate calculation. We show in Fig. 8.22 two metal blocks in identical varying fields. The first

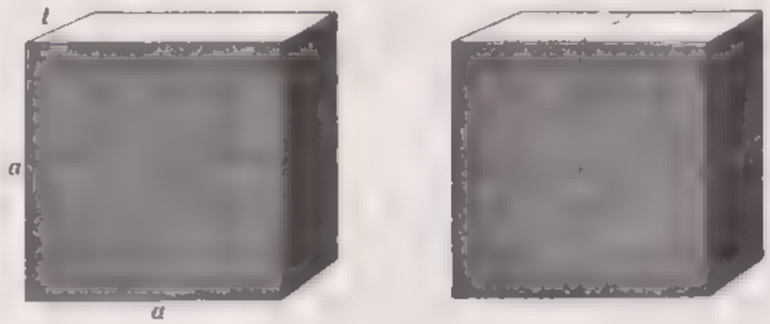


***Fig. 8.22** Solid and quartered block for eddy-current calculation.*

block is solid, while the second is divided into four insulated quarters. We calculate the eddy-current losses in the first block of dimensions $a \times a$ with thickness t and compare these with the losses in four quarter-blocks of area $\frac{a}{2} \times \frac{a}{2}$.

Figure 8.23 shows the first block broken into a set of thin square paths of length l and cross-section areas $t\,dx$. This is not

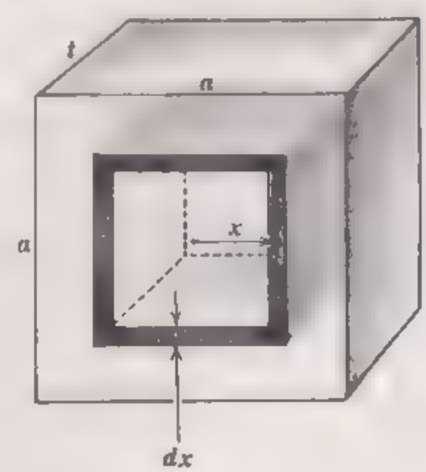


***Fig. 8.23** Idealized current path for calculation of eddy-current losses.*

the exact path that the current will take, but it provides a good first approximation. If we let $2x$ be the length of the side of the square path, its length l is $8x$. The area enclosed is $(2x)^2$. Now if B is directed into the face of the block and dB/dt is uniform over the block, the induced emf is proportional to the area enclosed by the path. The emf $\mathcal{E}$ around the path can thus be written as

$k(2x)^2$, where $k = dB/dt$. The average induced electric field E along the path is then $E = \dfrac{\mathcal{E}}{l} = \dfrac{4kx^2}{8x} = \dfrac{1}{2}kx$.

We now express the power dissipation due to the current around the path. We start with the general expression for power dissipation in a conductor, $P = Vi$, and transform it as follows:

$$P = Vi = Eli = ElAj = Ejv = \sigma E^2 v \tag{8.45}$$

where A = cross-section area of current path
v = volume of current path
σ = conductivity.

For the power dissipation in the path we have chosen we can then write

$$dP = \sigma E^2\, dv$$

Integration over all such paths gives the total rate of energy dissipation in the block.

$$P = \sigma \int E^2\, dv = \sigma \int_0^a \left(\frac{1}{2}kx\right)^2 t\, 8x\, dx = \frac{\sigma t k^2}{2} a^4$$

This result shows immediately the advantage of dividing the area into small islands, since it shows that the area of the block enters as its square. Thus if the block were divided into quarters, the power dissipation would be

$$P = 4\,\frac{\sigma t k^2}{2}\left(\frac{a}{2}\right)^4 = \frac{1}{4}\,\frac{\sigma t k^2}{2}\,a^4$$

or one-fourth the dissipation in the unquartered block. It is in planes perpendicular to the lines of B that the conducting path must be broken up. Thus the cores of transformers are usually built up of many laminas insulated from each other. Magnetic flux circulates around the laminas, but induced currents caused by $d\Phi/dt$ are held to a minimum in the plane perpendicular to the laminas by their small, thin cross section.

In the simplified example we have assumed that the resistivity of the conductor is high enough that eddy currents in the several elements do not seriously affect dB/dt in the material. This is realistic for many situations. At high enough frequencies the eddy currents indeed reduce the value of dB/dt inside the metal so as to limit the penetration of electromagnetic waves into a conductor.

This gives rise to the *skin effect* in conductors. For example, high-frequency currents along a wire are confined to the outer surface of the wire, called the *skin depth*, by this eddy-current effect. It is also this effect which makes metals opaque to light, which is itself an electromagnetic wave.

PROBLEMS

8.1 A metal rod of length L moves with a velocity v in a direction perpendicular to its axis and to a constant magnetic induction field B, as shown in Fig. P8.1*a*.

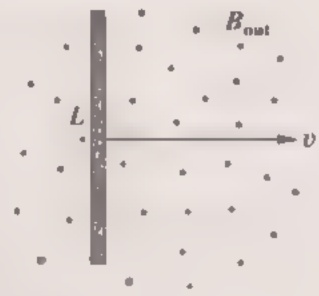


Fig. P8.1a

a Write the expression for the force on charges in the rod by virtue of the motion of the rod.

b What will be the magnitude and direction of the electric field set up by the separation of charges due to this motion?

c What will be the potential difference between the ends of the rod?

d If the rod moves on a stationary conducting loop as shown in Fig. P8.1*b*, what current will flow in the loop if the resistance of the loop and rod is R?

e Calculate the emf induced by the motion by means of Faraday's law of induction and compare the result with (*c*).

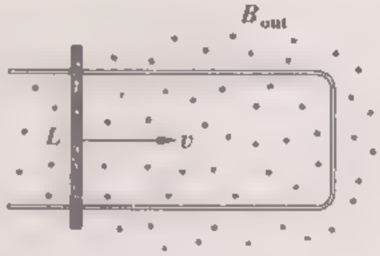


Fig. P8.1b

8.2 A coil of n turns and area A rotates with a frequency ω about a diameter that is perpendicular to a uniform magnetic induction field B. Calculate the peak emf (amplitude) induced in the coil.

8.3 Two identical coils are connected in series and spaced in such a way that one-half the flux from one coil threads the second coil. If the self-inductance of one coil is L henrys, find the self-inductance of the

pair of coils connected in series, assuming the coils are connected in such a way that the fluxes add (rather than subtract).

8.4 A small loop of wire of radius a is coaxial with a large loop of wire of radius b. The two loops are separated by the distance l. Under the assumption that $b \gg a$, compute the mutual induction between the loops. (*Note*: The problem is easy or hard depending on which loop you let produce the magnetic field.)

8.5 A conducting rod of length L rotates with an angular velocity ω about one end in a plane perpendicular to a uniform magnetic induction field B. Find the emf induced in the rod, first in terms of the motional emf and then by means of the Faraday induction law.

8.6 State a situation in which the emf in a conducting loop can be calculated by means of Faraday's induction law but cannot be obtained by means of motional emf.

8.7 A stiff wire bent into a semicircle of radius R, as shown in Fig. P8.7, is rotated at a frequency f in a uniform magnetic field B, as shown below. What are the amplitude and frequency of the induced voltage and of the induced current when the internal resistance of the meter is 1,000 ohms and the remainder of the circuit has negligible resistance?

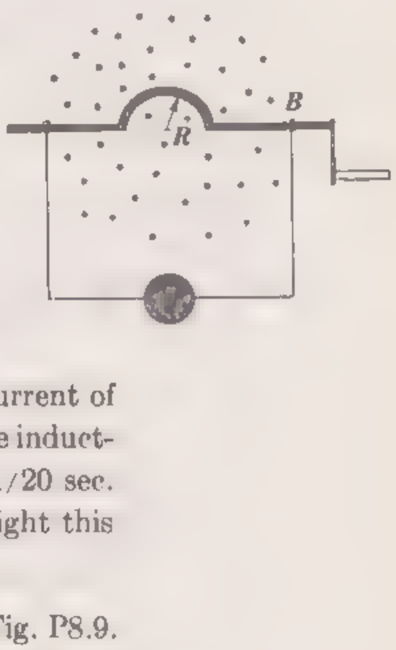


Fig. P8.7

8.8 A large electromagnet has an inductance of 5 henrys. If a current of 10 amp flows through its coils, what is the energy stored in the inductance? When the current is interrupted, it drops to 1 amp in 1/20 sec. Approximately what voltage is induced in the coil? How might this voltage have a detrimental effect on the coil windings?

8.9 A solenoid is moving toward a conducting loop as shown in Fig. P8.9. As viewed from the solenoid, what will be the direction of the induced current in the loop?

Fig. P8.9

8.10 The accelerating action in a betatron depends on the induced emf produced by a changing magnetic field. Electrons in a vacuum travel in circular paths in a magnetic field that increases with time. The radius of curvature of the electron orbit depends on the field B_R, *at* the orbit and perpendicular to its plane. The accelerating force depends on $d\Phi/dt$ over the area enclosed by the orbit.

a Show that for the radius of the orbit to stay constant as electrons are accelerated, the average field $\bar{B}$ within the orbit must be $2B_R$.

b Express the energy gained per cycle in terms of dB_R/dt.

8.11 The rectangle shown in Fig. P8.11 is moving with a uniform velocity V away from the long wire carrying current i. The wire and rectangle remain coplanar. Calculate the electromotive force induced in the rectangle. What is the mutual inductance between the two circuits as a function of time if the nearer wire of the rectangle is distance l from the long wire at time $t = 0$?

Fig. P8.11

8.12 A current loop is made by connecting the ends of two long parallel wires of radius a separated by a distance d between centers. Neglecting end effects and the magnetic flux within the wires, show that the self-inductance of a length l of the parallel wires is

$$L = \frac{\mu_0 l}{\pi} \ln \frac{d - a}{a}$$

8.13 The self-inductance of a coil can be computed by equating the stored energy in the magnetic induction field of the coil to the work necessary to build up the current in the coil from zero. Do this for the simple case of a toroidal solenoid of length l, cross section A, having N turns.

8.14 How can 50 volts be generated across the terminals on a 2-henry inductor of negligible resistance?

8.15 The inductance of a close-packed coil is 10 mh. The coil has 100 turns. Find the total magnetic flux through the coil when the current is 0.5 ma.

8.16 Find the magnetic energy stored *inside* a 1 m length of wire carrying 10 amp. The wire is 1 mm in radius and the current density is uniform.

8.17 A current of 10 amp produces a total flux of 10 webers in a closely wound coil of 200 turns. Find the energy stored in the magnetic field.

8.18 A horizontal magnetic field of strength B is produced across a narrow gap between square iron pole pieces as shown in Fig. P8.18. A closed square wire loop l meters on a side is allowed to fall with the top of the loop in the field. The loop has a resistance R and a weight mg. Find the terminal velocity of the loop while it is between the poles of the magnet. How would this velocity be modified if the cross section of the wire of the loop were doubled?

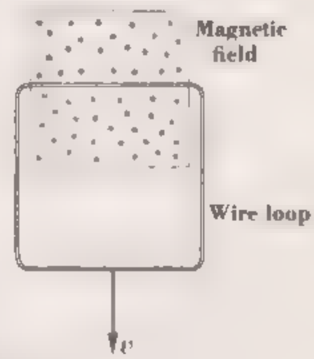


Fig. P8.18

8.19 A uniform field of induction B is changing in magnitude at a constant rate dB/dt. Given a mass m of copper to be drawn into a wire of radius r and formed into a circular loop of radius R, show that the induced current in the loop does not depend on the size of the wire or the loop and is given by

$$i = \frac{m}{4\pi\rho\delta}\frac{dB}{dt}$$

where ρ is the resistivity and δ is the density of copper.

NINE

Magnetism in Matter

9.1 Introduction

In this chapter we show what modifications must be made in the picture of magnetism in order to account for magnetic effects in matter. After a brief preliminary discussion of the sources of magnetic behavior in matter, we derive some general relationships between the magnetic parameters in common use. Following this, we discuss in more detail the atomic contributions to magnetism, which account for the three most important types of magnetic behavior: *paramagnetism*, *diamagnetism*, and *ferromagnetism*. We next study some problems relating to the *magnetization* of magnetic bodies placed in a uniform field and then discuss briefly the various ways of considering permanently magnetized bodies. Finally, we consider briefly the *magnetic circuit*, of particular importance in the design of electromagnets.

9.2 Magnetic Contributions of Matter

We begin our discussion by referring to a very simple experimental arrangement that can be used to measure the magnetic contribution of matter. Figure 9.1 shows a toroidal solenoid, with a small

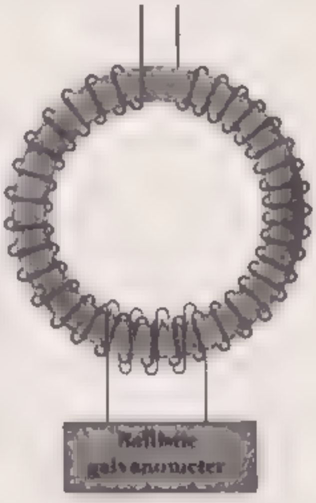


Fig. 9.1 Toroidal solenoid with secondary winding for flux measurement.

extra or secondary winding that can be connected to a ballistic galvanometer. With the space inside the toroid empty, the ballistic galvanometer can be used to measure the total flux $\int \mathbf{B} \cdot d\mathbf{A}$ inside the coil due to a current i in the toroid by measuring the galvanometer swing when the current is suddenly turned off. When the experiment is repeated with matter filling the entire toroid, a different deflection is found, in general, for the same current. It is this extra contribution to the total flux that we wish to describe.

In a vacuum the magnetic induction in the solenoid is given by

$$B = \mu_0 \frac{Ni}{L} = \mu_0 j^s \qquad \text{webers/m}^2 \tag{6.10}$$

where j^s is the solenoidal current density as defined in Sec. 6.3. This value of **B** was obtained by integrating

$$d\mathbf{B} = \frac{\mu_0}{4\pi} \frac{i\, d\mathbf{l} \times \hat{\mathbf{r}}}{r^2} \tag{6.2}$$

for the geometry of the solenoid or, alternatively, by using Ampère's circuital law,

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 i$$

around the path inside a toroidal solenoid.

Experiments with matter filling the toroid show that, in general, **B** is modified by the presence of matter. With matter filling the toroidal solenoid we must add another term to the flux $\Phi = BA$. This extra term increases **B** (or in some cases, decreases it). We thus rewrite Eq. (6.10) as

$$B = \mu_0(\mathscr{j}_{\text{free}} + \mathscr{j}_{mag}) \tag{9.1}$$

where $\mathscr{j}_{\text{free}} = Ni/L$ is the solenoidal current density in the coil, and $\mathscr{j}_{mag}$ is an *effective* solenoidal current density that describes the magnetic effect of the matter.

It is possible to show that if magnetized matter consists of a certain number of magnetic dipoles aligned in the direction of the magnetic induction field, these magnetic dipoles will indeed produce a macroscopic magnetic induction field of the same kind as would be obtained from this effective solenoidal current density. We shall show this connection after first describing qualitatively the nature of magnetic materials.

The magnetic effects in magnetic materials are due to atomic magnetic dipoles in the materials. These magnetic dipoles result from the effective current loops of electrons in atomic orbits, from effects of electron "spin" (intrinsic magnetic moments of electrons, which can be described in terms of an effective toplike motion of the electrons), and from the magnetic moments of atomic nuclei. The effect of an external field is to tend to align these magnetic moments along the field direction. We may characterize the degree of alignment as the total magnetic moment per unit volume of the material, where the total magnetic moment is the vector sum of the individual moments. This is the magnetization **M** of the material. This quantity **M** plays a role in magnetism similar to that of the polarization **P** in dielectrics.

In order to show the connection between **M** and the effective solenoidal current density $\mathscr{j}_{mag}$, we consider a small cube (see Fig 9.2) which contains a certain number of atomic magnetic dipoles aligned in the direction of the external magnetic field. More correctly, since not all dipoles will be pointed in the field direction, these dipoles represent the *net* magnetization of the cube. As we have seen, the magnetic moment is a vector quantity, and the field of each dipole is given by Eq. (6.24). We want to find the magnetic induction field at points away from the cube such as the point S. We argue that the field at S could be equally well represented by a

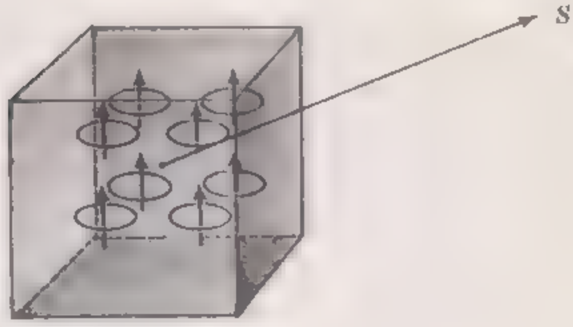


Fig. 9.2 Aligned dipoles represent the net magnetization of a small cube. They produce a magnetic induction field at points like S.

sheet of current i_{mag} circulating around the sides of the cube as shown in Fig. 9.3. Such a current loop acts as a dipole, just as do the individual atomic dipoles, so from any point S away from the cube, the field would be the same, whether due to individual

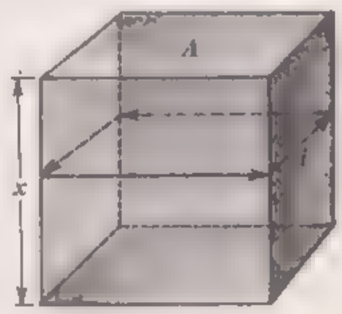


Fig. 9.3 Effects of magnetic dipoles at external points can be simulated by an effective current i_{mag} circulating around each volume element.

atomic dipoles or to a current around the cube. The dipole moment of the cube is

$$p_m = i_{mag}A$$

where A is the area of the loop. We may also express the magnetization M, or magnetic moment per unit volume in the cube. This is

$$M = \frac{i_{mag}A}{xA} = \frac{i_{mag}}{x} = j^s_{mag} \qquad \text{amp/m} \tag{9.2}$$

where j^s_{mag} is the equivalent solenoidal current density circulating around the cube and giving the same magnetic moment as the sum of the individual atomic magnetic dipoles.

We now extend this argument to a large cylinder of magnetic material. In order to avoid end effects, let us consider a toroidal cylinder placed inside the toroidal coil of Fig. 9.1. We examine a small slice of this sample, as shown in Fig. 9.4. Consider the slice to be made up of many identical cubes such as the one described above. As before, we replace the magnetic effect of the individual magnetic dipoles in each cube by a circulating solenoidal current

density j^s_{mag}. The effects of all the current loops will cancel inside the slab, since currents on the surfaces of neighboring cubes are equal and opposite, leaving only the circulating currents around the outer surface of the slab. The magnetic moment per unit volume in the slab is that of the individual cubes, that is, $M = j^s_{mag}$. But now we may consider j^s_{mag} as the effective current density circulating around the outer surface of the slab. This is the *Amperian surface current*, which may be used to replace the magnetic effects of the actual atomic magnetic moments. Since the magnetic induction

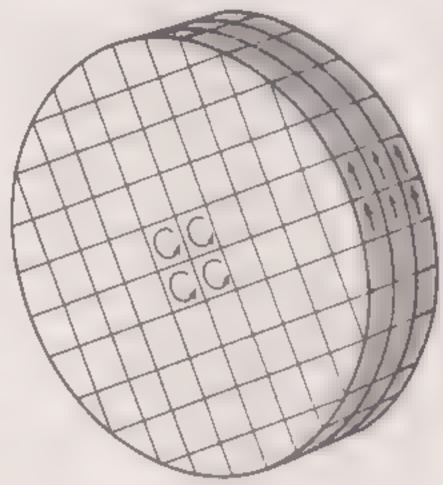


***Fig. 9.4** The circulating currents that account for the magnetic properties of matter can be replaced by an equivalent solenoidal current density called the Amperian surface current.*

field intensity inside a long solenoid is uniform, the Amperian current gives a uniform field inside. This uniform field corresponds to the *average* field inside the magnetic material. The Amperian current in magnetic materials corresponds to the induced surface charges in dielectric materials. We have replaced the atomic currents that produce magnetic effects in matter by a fictitious surface current that is much easier to handle mathematically.

The geometry we have chosen for this discussion is very special. In effect, the toroidal solenoid is a long rod without any ends. If we had placed a smaller piece of matter within the coil, having finite ends, the problem would have been much more complicated, since in order to satisfy boundary conditions, the original vacuum field would be seriously modified by the sample. We show in Sec. 9.12 that whenever the magnetization M of a medium has a component perpendicular to the surface of the medium, the original field outside the medium is modified by its presence. In our special case, the contribution of the matter is a field that exists only in the matter and is everywhere parallel to the original vacuum field of the solenoid. The general problem of finding the field inside a

piece of matter of any shape is more difficult, and we treat only some simple cases later on.

An additional complication that arises is that if the magnetization of matter is not uniform, the individual current loops will not cancel completely, and we must add to the effective surface current density a term involving a volume distribution of effective currents. For the time being, we shall continue to consider the very much simpler case of a toroidal sample, uniformly magnetized.

In later sections we shall see that certain kinds of dipoles tend to be aligned on the average in the direction of the field (paramagnetism), while other effects result in induced dipoles aligned against the field (diamagnetism).

9.3 Magnetic Field Intensity H

In Eq. (9.1) we have seen that the magnetic induction in the sample arises from two sources, one the current in the coil—the external effect—and the other the magnetization of the sample, which we find can be calculated from an effective current density. The term that depends on external effects is of considerable importance, and we shall call it the magnetic field intensity **H**. In the case of the solenoid

$$H = \frac{Ni}{L} = j^s_{\text{free}} \qquad \text{amp/m}$$

If the solenoid is empty, there is no contribution from the magnetization, and

$$\mathbf{B} = \mu_0 \mathbf{H}$$

that is, **B** and **H** differ only by a constant factor μ_0. **H** represents a vector field somewhat similar to the induction field **B**, but differing in that it ignores the effect of matter, much as **D** in electrostatics ignores the effects of a dielectric. Later, in Sec. 9.12, we shall see that for non-toroidal samples, **H** is affected by boundary effects in magnetic materials. Thus our present definition of $\mathbf{H} = Ni/L$ is incomplete. The importance of the **H** field lies partly in its convenience in the solution of problems involving boundaries. Equally important is the fact that magnetization is induced in matter to an extent that depends on **H**. It is very often true that the induced magnetization **M** is *proportional* to **H**, and we can then describe

the properties of the material in terms of its *magnetic susceptibility* χ_m or its *magnetic permeability* μ. We define these terms in the next section.

9.4 Magnetic Parameters in Matter

For the special case of a toroidal sample, Eq. (9.1) can now be rewritten as

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) \tag{9.5}$$

Here the first term is the contribution of the current in the solenoid, and the second term is due to the magnetization of the sample and can be simulated by an effective solenoidal surface current. This equation is true in general, but for other than toroidal sample shapes, the calculation of **H** becomes more involved. We discuss this problem in Sec. 9.12.

For the large class of materials in which the magnetization is proportional to **H**, we may define the magnetic susceptibility of the material by χ_m, where

$$\mathbf{M} = \chi_m \mathbf{H} \tag{9.6}$$

As defined here, χ_m is a pure number, since **M** and **H** have the same units. It then follows from Eq. (9.5) that

$$\mathbf{B} = \mu_0(1 + \chi_m)\mathbf{H} \tag{9.7}$$

If we write

$$\mu = \mu_0(1 + \chi_m) \tag{9.8}$$

where μ is the magnetic permeability, we have

$$\mathbf{B} = \mu \mathbf{H} \qquad \text{or} \qquad \vec{B} = \mu \vec{j}_{\text{free}} \tag{9.9}$$

It is useful to define a *relative permeability* μ/μ_0. The relative permeability in the mks system has the same numerical value as μ, the permeability, in cgs magnetic units (emu).

We can classify magnetic matter in terms of the susceptibility χ_m. In paramagnetic materials, χ_m is positive and much less than 1. In diamagnetic materials, χ_m is negative and also small compared with 1. The characteristic of ferromagnetic materials is that χ_m is positive and very large. However, in ferromagnetic materials **M** is not accurately proportional to **H**, so χ_m is not a constant,

except over a small range of **H**. We discuss these three cases in detail later.

We may also classify magnetic materials in terms of the magnetic permeability μ. From its definition, we may characterize magnetic matter as follows:

Diamagnetic: $\mu < \mu_0$

Paramagnetic: $\mu > \mu_0$

Ferromagnetic: $\mu \gg \mu_0$

9.5 *Characteristics of* B *and* H

We now turn back to our earlier consideration of the two magnetic vector fields: **B**, the magnetic induction, and **H**, the magnetic intensity. In the earlier discussion of **B**, in Secs. 6.2 and 6.4, we found that in a vacuum the lines of **B** are continuous and form closed loops. This will continue to be true in the presence of matter, since we have shown that the magnetic effect of uniformly magnetized matter can be simulated by an effective current density circulating around the magnetic material. Thus, we imagine replacing the matter by suitable coils of appropriate shape and current density and thereby return effectively to the vacuum case. Therefore, under all situations the surface integral of **B** over a surface completely enclosing a volume is zero. Even when matter is present, we get

$$\int_{CS} \mathbf{B}\cdot d\mathbf{S} = 0 \tag{9.10}$$

That is, the number of lines of **B** emerging from any volume surrounded by a surface S is the same as the number entering it (since all lines are continuous and have no beginnings or endings). This is Gauss' law applied to **B** and is exactly the same as the rule for **E** in electrostatics in a region where there are no charges to act as sources or sinks for the lines of **E**. Under no conditions are there sources or sinks for lines of **B**.

Another property of lines of **B** refers to the line integral around a closed path. We showed in Sec. 6.4 that

$$\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0 i \tag{6.14}$$

where i is the current threading the closed path taken. Now that we are including the effects of magnetic matter, we should rewrite this as

$$\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0(i_{\text{free}} + i_{mag}) = \mu_0 i_{\text{total}} \tag{9.11}$$

where i_{free} is the real current and i_{mag} is the fictitious Amperian current that describes the magnetic behavior of matter.

This last equation for the line integral of **B** can be made more useful if we derive from it an equation for **H**. The first step is to use the general equation (9.5), $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$, to express **B** in terms of **H** and **M**. Making this substitution, we get

$$\frac{1}{\mu_0}\oint \mathbf{B}\cdot d\mathbf{l} = \oint \mathbf{H}\cdot d\mathbf{l} + \oint \mathbf{M}\cdot d\mathbf{l} = i_{\text{total}} \tag{9.12}$$

Now if we can evaluate the term involving **M**, we can solve for $\oint \mathbf{H}\cdot d\mathbf{l}$. But since we already have $\mathbf{M} = j^s_{mag}$, we see at once that $\oint \mathbf{M}\cdot d\mathbf{l} = i_{mag}$. Let us clarify this point with the aid of Fig. 9.5, which shows a section of a toroidal solenoid sample on which the

Fig. 9.5 Evaluation of the line integral of the magnetization around a closed path.

Amperian surface current has been indicated. This current replaces the effects of the oriented magnetic dipoles of the sample. The magnetization is produced by real current in a solenoidal coil surrounding the sample and not shown in the figure. We evaluate the line integral $\oint \mathbf{M}\cdot d\mathbf{l}$ around the closed path *abcd*. Along *ab*, M is parallel to the path and equal to j^s_{mag}. If we let the distance *ab* be x, the integral from a to b becomes $j^s_{mag}x$. The line integral is zero over the rest of the path, since M is perpendicular to the paths *bc* and *da* and is zero over the path *cd*. Since $j^s_{mag}x$ is the total Amperian current i_{mag} enclosed by the path, we have

$$\oint \mathbf{M}\cdot d\mathbf{l} = i_{mag}$$

If we substitute this in Eq. (9.12), noting that $i_{\text{total}} = i_{\text{free}} + i_{mag}$, we find

$$\oint \mathbf{H} \cdot d\mathbf{l} = i_{\text{free}} \tag{9.13}$$

This is a fundamental property of **H** and is the analogue of Eq. (6.15), $\oint \mathbf{B} \cdot d\mathbf{l} = 0$, which we found for the vacuum case. Equations (9.10) and (9.13) are basic equations for **B** and **H** that are valid under all conditions.

In any region of space where there are no current-carrying conductors (and no moving charges), i_{free} is zero, and Eq. (9.13) becomes

$$\oint \mathbf{H} \cdot d\mathbf{l} = 0 \tag{9.14}$$

9.6 *Boundary Conditions for* B *and* H

We shall now use these powerful new rules to show how **B** and **H** behave at the boundary of a magnetic material. The procedure will recall the similar development of the conditions on **E** and **D** at a dielectric boundary, since the mathematics is identical.

In the case of **B**, we proceed by imagining a cylindrical Gaussian surface as shown in Fig. 9.6, drawn so as to have one surface in

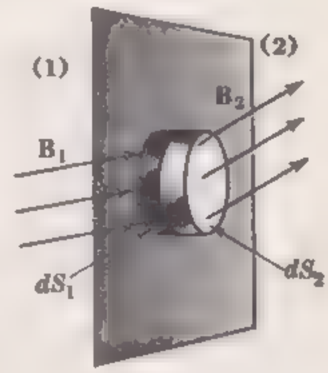


***Fig. 9.6** Gaussian surface for obtaining boundary condition on B between two different magnetic media.*

region 1 and the other in region 2. The two regions have different magnetic properties; for example, region 1 may be a vacuum, while region 2 could be a magnetic material. We make the thickness of the cylinder vanishingly small, so the lines of **B** that come out of

the sides of the cylinder may be neglected. Then we can write, from Eq. (9.10),

$$\int_{S_1} \mathbf{B}_1 \cdot d\mathbf{S}_1 = -\int_{S_2} \mathbf{B}_2 \cdot d\mathbf{S}_2 \tag{9.15}$$

This means that the normal components of $\mathbf{B}_1$ and $\mathbf{B}_2$ are equal, or

$$B_{n1} = B_{n2} \tag{9.16}$$

We next study the behavior of $\mathbf{H}$ at this boundary by taking the line integral as shown in Fig. 9.7. If the magnetic matter is not

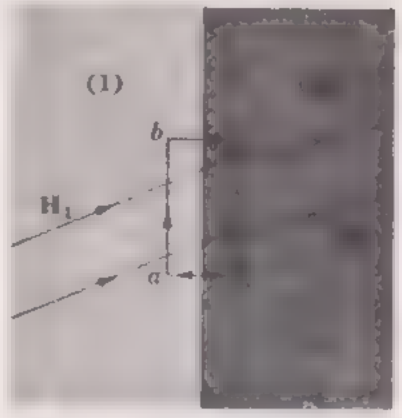


***Fig. 9.7** Line integral for obtaining boundary condition on H between two different magnetic media.*

wrapped with a coil carrying current, as we are to assume, we can use Eq. (9.14) to show

$$\int_a^b \mathbf{H}_1 \cdot d\mathbf{l}_1 = -\int_c^d \mathbf{H}_2 \cdot d\mathbf{l}_2 \tag{9.17}$$

This means that the tangential components of $\mathbf{H}$ are equal on the two sides of the boundary, or

$$H_{t1} = H_{t2} \tag{9.18}$$

So far we have made no assumption as to how $\mathbf{H}$ and $\mathbf{B}$ are related. When we are dealing with materials in which $\mathbf{B} = \mu\mathbf{H}$, we can determine the behavior of $\mathbf{B}$ and $\mathbf{H}$ at a boundary, using the rules just derived for B_n and H_t. Since these conditions are identical with those for D_n and E_t found for dielectric boundaries, we may take over the earlier results directly, to write

$$\frac{\tan \Phi_1}{\tan \Phi_2} = \frac{\mu_1}{\mu_2} \tag{9.19}$$

for the directions of **B** and **H** on the two sides of a boundary (see Sec. 5.6 for the electrostatic case).

9.7 Stored Energy in Magnetic Matter

We have already found in Sec. 8.11 that the energy density in a vacuum containing a magnetic induction field **B** is

$$\frac{U}{\text{vol}} = \frac{1}{2}\frac{B^2}{\mu_0} \qquad \text{joules/m}^3 \tag{8.32}$$

Identical arguments, using a toroidal solenoid filled with magnetic material, allow the calculation of energy density in magnetized matter. We need only replace the vacuum equation

$$B = \mu_0 \mathcal{J}_{\text{free}} \tag{6.12}$$

by

$$B = \mu \mathcal{J}_{\text{free}} \tag{9.20}$$

to obtain the required result. It is left as a problem for the student to show by this means that the energy density in matter is given by

$$\frac{W}{\text{vol}} = \frac{1}{2}\frac{B^2}{\mu} = \frac{1}{2}BH = \frac{1}{2}\mu H^2 \tag{9.21}$$

In Sec. 5.8 we showed how the stored energy in the electric field could be used to calculate the forces acting on a dielectric plate placed between capacitor plates. Similar problems exist in the case of magnetic materials. In the dielectric case, if the system is isolated, mechanical work done by the electric field in displacing the dielectric is done at the expense of stored electrostatic energy. However, in the problem discussed, a battery was connected to the capacitor plates to maintain constant electric field. We showed that the work done by the battery during a displacement of the dielectric was twice the mechanical work done by the electric forces. As a result, instead of finding a *decrease* in stored energy of an amount $-dU$, we found that the stored energy was *increased* by an amount $-dU + 2\,dU = dU$. The work-energy relation then becomes $F\,dx = dU$ instead of $F\,dx = -dU$, where $F\,dx$ is positive work done by the electric field. This result is general for any linear system (for which the susceptibility is constant).

In the case of a magnetic system, such as in Figures P.9.2 and P.9.9, the same effect is seen. Motion of magnetized material causes work to be done by the external source which has built up the stored energy in the magnetic field, and for linear systems the work done by the external energy source compensates twice over for the loss of stored magnetic energy. $F\,dx = dU$ should thus be used in the problems cited.

9.8 Paramagnetism

The essential feature of a paramagnetic material is that it has a positive but small magnetic susceptibility ($\chi_m \ll 1$), which results from the existence of permanent magnetic dipoles that are free to be oriented under the influence of an external field. The dipoles exert very small forces on each other; therefore, at ordinary temperatures, thermal vibrations of the solid ensure random orientation of the dipoles, giving an average magnetization of zero. The greater the applied magnetic field, the more nearly the dipoles tend to be aligned and the greater is the net magnetic moment per unit volume, or magnetization **M**.

Let us consider the magnetic susceptibility of a paramagnetic substance, considering that there are N atoms in a unit volume, each having a magnetic dipole p_m, which can be aligned in an applied field. Let us assume, as is often the case, that in an applied field there are only two possibilities—either the dipoles are aligned parallel to the field or they are aligned antiparallel.[1] The magnetization at a given temperature and applied field depends on the excess number of magnetic moments aligned parallel to the field over those aligned antiparallel. The excess fraction of moments aligned parallel to the field is given by

$$f = \frac{p_m B}{3kT} \tag{9.22}$$

where k is the Boltzmann constant and T the absolute temperature. We shall not derive this result but we note that the numerator relates to the energy of orientation of dipoles in the external field,

[1] It is a result of quantum mechanics that magnetic dipoles in a field are subject to the condition that they may take up only certain specified average positions with respect to the field direction (see Sec. 14.7). If all orientations are allowed, the same result is obtained within a numerical constant.

and the denominator relates to the thermal energy of the substance. Thus the higher the temperature, the more seriously thermal vibrations interfere with alignment and the smaller is the excess fraction of aligned moments.

The susceptibility is obtained from the last equation as follows: The magnetization is given by [1]

$$M = Np_mf = \frac{Np_m{}^2\mu_0H}{3kT}$$

Then

$$\chi = \frac{M}{H} = \frac{N\mu_0p_m{}^2}{3kT} \tag{9.23}$$

This dependence of the susceptibility of paramagnetic materials on T^{-1} is known as the Curie law.

Figure 9.8*a* shows a plot of the magnetization M against the magnetizing field H. The slope of this curve is the susceptibility χ. A plot of the susceptibility χ against T, in Fig. 9.8*b*, shows the way the susceptibility falls off with temperature, according to Eq. (9.23). It is common practice to plot the reciprocal $1/\chi$ against temperature, in which case a straight-line plot as shown in Fig. 9.8*c* is obtained.

We now want to investigate the nature of p_m, the atomic dipole moment. Beginning with a discussion of free atoms or ions, we avoid the use of quantum mechanics by accepting certain simple facts about atomic structure that result from their quantum-mechanical nature. Classical considerations will then answer our questions. Thus we start with an atomic model in which electrons revolve in orbits about the much heavier nucleus. These electrons in motion act as small current loops and therefore contribute magnetic moments $p_m = Ai$, where A is the area of the orbit. We begin by calculating p_m for a simple orbit, such as that of a single electron around a nucleus in a free atom.

The area of a circular orbit is $A = \pi r^2$, and the current due to an electron of charge e is e times the number of times the electron passes any point in the orbit per second, or $i = ef = ev/2\pi r$, where f is the frequency of rotation and v the electron velocity. We determine v classically in terms of the orbit radius by equating the

[1] The magnetic induction field B has been replaced here by μ_0H. This amounts to neglecting the effect of M in the equation $B = \mu_0(H + M)$, as justified by the small value of M in paramagnetic materials.

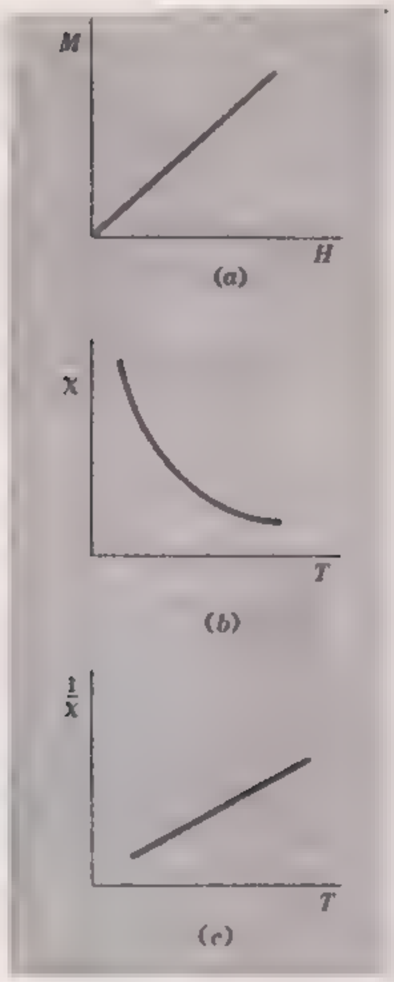


Fig. 9.8 Paramagnetic susceptibility. (a) Slope of magnetization M versus applied field H curve gives the susceptibility $\chi = M/H$. (b) Susceptibility χ as a function of temperature. (c) $1/\chi$ versus T plot is a straight line.

Coulomb force between the electron and the nucleus of charge Ze to the centripetal force holding the electron in orbit. Thus we write

$$ma = \frac{mv^2}{r} = \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r^2} \tag{9.24}$$

to give

$$v = e\left(\frac{Z}{4\pi\epsilon_0 mr}\right)^{1/2}$$

Substitution gives for the moment,

$$p_m = Ai = \frac{e^2}{2}\left(\frac{Zr}{4\pi\epsilon_0 m}\right)^{1/2} \tag{9.25}$$

For an atomic radius of 10^{-10} m this gives $p_m \sim 10^{-23}$ amp-m^2. Most of the electrons circulating about the nucleus need not be considered as contributors to the permanent magnetic moment. This results from the tendency in atoms for the electrons to pair off in such a way that they produce equal and opposite moments

that cancel out. Classically this would result from two electrons going around in opposite directions in the same orbit. However, in many atoms and ions there is one electron (or more) whose moment is not canceled by that of another.

Another source of magnetic moment, in addition to the contributions from orbital motion described above, is the spin of the electron. That is, the electron behaves as though it were spinning on its own axis. This again produces an effective current loop with a consequent magnetic moment. In atoms and ions it is again usually true that electrons have a strong tendency to pair up in arrangements having opposite spins, with a consequent cancellation of the magnetic moments. However, all atoms or ions having an odd number of electrons, as well as some other cases of incomplete cancellation, will exhibit spin magnetism. The magnetic moment due to electron spin is of the same order of magnitude as that due to orbital motion.

As a matter of fact, in solids the magnetic moment due to electron spin is usually the principal contributor to the magnetic properties. This results from the effect of the local electric fields that exist in solids on the orbital motion of electrons. These fields tend to cause the plane of the orbit to precess rapidly so that the average value of the magnetic moment in any given direction is zero. Thus the magnetic moment of the free atom or ion is usually not the correct value in solids or liquids.

An additional, though much smaller, contribution to the magnetization in matter comes from the fact that some nuclei have magnetic moments. This can be thought of as due to the motion of charge within the nucleus. The magnitude of nuclear moments is about 10^{-3} that of electronic moments.

Substituting real values of moments in Eq. (9.23), we find that paramagnetic materials at room temperature have at most a susceptibility of order 10^{-3}. Thus magnetic effects in these materials are quite weak, and the value of μ differs from μ_0 by less than 1 per cent. Even so, it is possible to measure these effects with considerable accuracy. One method is to use the arrangement shown in Fig. 9.1, though a more accurate result can be obtained by measuring the force on a paramagnetic sample in a nonuniform magnetic field.

9.9 Diamagnetism

Diamagnetism results from the negative magnetic moments induced in all matter upon the application of an external magnetic field. It is characterized by a *negative* susceptibility and is independent of temperature. Since it is small compared to paramagnetic (and ferromagnetic) magnetization, it can be observed directly only in materials that are otherwise nonmagnetic. This negative susceptibility can be understood on the basis of Faraday induction acting on the orbital motions of electrons in atoms or ions.

We can understand this effect and its negative sign by thinking of an electron in a circular orbit, such as shown in Fig. 9.9, and

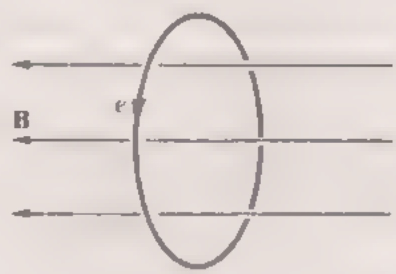


***Fig. 9.9** Effect of external field on electronic orbit.*

asking what happens if a field **B** is introduced perpendicular to the plane of the orbit. According to the Faraday law, there is an emf acting on the electron given by

$$\mathcal{E} = -\frac{d\Phi}{dt} = -A\frac{dB}{dt}$$

during the time the field is changing, where A is the area of the orbit. This emf has the effect of changing the nature of the circulating motion of electrons in atomic orbits. We show below that for magnetic induction fields that make reasonably small changes in the circular motion, the size of the orbit is not altered. We therefore make a calculation based on constant orbital radius. The emf produced by the changing **B** field results in an electric field E_t, which acts tangentially to the direction of the electronic motion. The magnitude of this field is

$$E_t = \frac{\mathcal{E}}{2\pi r}$$

As a result of this field, the electron will be accelerated according to Newton's law, giving

$$m\frac{dv}{dt} = eE_t = \frac{e\mathcal{E}}{2\pi r} = -\frac{e}{2\pi r}\frac{d\Phi}{dt} \tag{9.26}$$

To find the total change in velocity during the time the magnetic flux is being increased, we integrate both sides of Eq. (9.26) to get

$$\Delta v = -\frac{e}{2\pi rm}\Phi$$

where Φ is the final magnetic flux through the loop and Δv is the total *change* in velocity due to the force from the Faraday induction. When we express this as a change in angular velocity through the relationship $\Delta v = r\omega_L$, we have

$$\omega_L = -\frac{e}{2\pi r^2 m}\Phi = -\frac{e}{2\pi r^2 m}AB = -\frac{e}{2m}B \tag{9.27}$$

This change in angular velocity, ω_L, is called the *Larmor frequency*. The negative sign means that ω_L is negative (clockwise) when **B** is positive (outward), in conformity with Lenz's principle.

This additional velocity accounts for an induced magnetic moment that always acts antiparallel to the field, to give a negative susceptibility. Before showing this we establish the fact that the radius of the orbital motion remains unchanged. We begin by showing that, in the absence of a magnetic induction field, the electrostatic force between nucleus and electron provides the centripetal force that holds the electron in circular motion. Thus we write

$$F = ma = \frac{mv^2}{r} = mr\omega_0^2 \tag{9.28}$$

where F is the electrostatic force and ω_0 is the angular velocity of the electron.

When a magnetic field B is turned on, the velocity ω_0 changes to $\omega_0 + \omega_L$, as we saw above (ω_L may be either positive or negative, depending on the direction of B with respect to the original motion). Also, a new term $e(\mathbf{v} \times \mathbf{B})$ is added to the centripetal force. For the geometry we have taken, this term is evB. We wish to show that this new magnetic force due to the applied B accounts for the change in the centripetal force necessary to balance the new rotational velocity $\omega_0 + \omega_L$. We do this by writing the equation for the balance of forces similar to Eq. (9.28) under the new conditions of an applied **B**. This is

$$F = mr(\omega_0 + \omega_L)^2 + evB \tag{9.29}$$

When this is multiplied out, using $v = r(\omega_0 + \omega_L)$, we find

$$F = mr\omega_0^2 + 2mr\omega_0\omega_L + mr\omega_L^2 + er\omega_0 B + er\omega_L B \tag{9.30}$$

When B is expressed in terms of ω_L according to Eq. (9.27), the second and fourth terms exactly cancel, leaving $F = mr\omega_0^2$ plus terms in ω_L^2, which can be neglected as long as $\omega_L \ll \omega_0$.[1] This is the case we wish to consider. This important result tells us that after the magnetic field is applied there is a cancellation of terms which shows that the original electrostatic force for an orbit of the original radius r is still the required value for equilibrium. The applied field **B**, which causes a change in velocity as it is turned on, also provides just the necessary additional (magnetic) force to keep the orbit in equilibrium. Thus the radius remains constant.

We next calculate the contribution to the magnetic moment of the orbit due to the additional angular frequency ω_L. The magnetic moment of a circular current is $p_m = iA$, where i is the current and A the area of the current loop. Since $\omega_L = 2\pi f_L$ and $i = ef_L$, where f_L is the frequency of rotation due to the magnetic field, we have

$$p_m = iA = ef_L A = \frac{e\omega_L}{2\pi} A \tag{9.31}$$

Lenz's rule gives the result that the direction of the induced dipole moment is always such as to minimize the change in flux threading the loop. This means that p_m always acts in opposition to the applied **B** change, resulting in the negative magnetic susceptibility of diamagnetism.

An example that brings out the nature of the diamagnetic effect is the case of two electrons traveling in opposite directions in the same orbit. In the absence of an external field, the magnetic effects of the orbital motion of these electrons will vanish. When a magnetic field is applied, however, the velocity of one electron will be increased and that of the other will be decreased by the Larmor frequency, so both orbital motions will contribute to a negatively oriented induced magnetic dipole.

We have examined only the case where the applied magnetic field is perpendicular to the orbit. Although we shall not show it

[1] The fifth term involves ω_L^2 since from Eq. (9.27), $B \propto \omega_L$.

here, the induced diamagnetic moment is the same for any orientation of the orbit with respect to the field. In the nonperpendicular case, however, ω_L will not be parallel to ω_0. That is, the motion in the applied field is circular motion in an orbit that itself precesses in the magnetic field with a precessional frequency ω_L.

Since the diamagnetic effect involves induced magnetic moments that are independent of the orientation of atoms, thermal vibrations do not affect diamagnetic susceptibilities. A plot of diamagnetic susceptibility against temperature is thus as shown in Fig. 9.10. We are neglecting certain other effects such as the

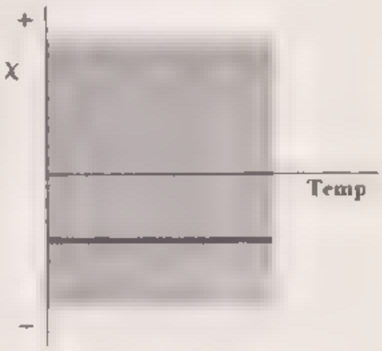


***Fig. 9.10** The diamagnetic susceptibility χ is negative and approximately independent of temperature.*

changing of lattice constants with temperature, which may make the diamagnetic susceptibility vary slightly with temperature.

Since paramagnetic susceptibilities decrease with temperature, while the diamagnetic term stays essentially constant, it follows that all materials become diamagnetic at high enough temperatures.

Table 9.1 gives a listing of the net susceptibility of a number of paramagnetic and diamagnetic materials. Ferromagnetic materials will be discussed in Sec. 9.10

Some metals are diamagnetic, while others are paramagnetic. In the case of metals, the net susceptibility is made up of contributions from the ion cores and the conduction electrons. There is a diamagnetic term contributed by the conduction electrons. This has to do with a quantum-mechanical effect in the presence of a magnetic field. There is also a paramagnetic contribution from the electron spins, which more than cancels the diamagnetic term. Only a small fraction of the electron spins contribute to the paramagnetism, since most of the conduction electrons are arranged in pairs with opposite spin orientation. Since the net susceptibility of a metal is made up of a number of terms, it is often difficult to predict whether a given metal will be diamagnetic or paramagnetic.

Table 9.1 Magnetic Susceptibilities, $\chi_m = M/H$, of Various Substances *

(At approximately room temperature)

Substance	$\chi_m \times 10^{-6}$, *mks units*
Aluminum	**+0.82**
Iron ammonium alum	**+38.2**
Calcium	**+1.4**
Chromium	**+4.5**
Cuprous oxide	**+1.5**
Ferric oxide	**+26.0**
Magnesium	**+0.69**
Manganese	**+1.**
O_2 liquid (−219°C)	**+390**
Platinum	**+1.65**
Tantalum	**+1.1**
Bismuth	**−1.7**
Cadmium	**−0.23**
Copper	**−0.11**
Germanium	**−0.15**
Helium	**−0.59**
Gold	**−0.19**
Lead	**−0.18**
Zinc	**−0.20**

* Paramagnetic materials have positive susceptibilities, diamagnetic materials have negative susceptibilities. These values are given in mks units and are 4π times values quoted for cgs units. This factor results from the cgs definition of magnetization, $B = H + 4\pi M$.

9.10 Ferromagnetism

We now come to the most easily observed of all magnetic effects, and indeed to the historic beginning of the study of magnetism: ferromagnetism, so called because of its occurrence in metallic iron and in a number of iron compounds. The experimental facts about ferromagnetic materials include the following. Although the magnetization is not usually proportional to **H**, as in the previously studied materials, in certain situations a susceptibility of several thousand can be measured, and very large magnetizations can be obtained. The value of the magnetization depends not only on the applied field but also on the previous history of the sample. In some cases, for example, a sample may retain its magnetization even in the absence of an external applied field. This is the source

of the permanent magnets with which we are all familiar. However, it is notable that the very same material that can show such a large permanent magnetization can also exist in a state showing little or no permanent magnetization. These phenomena alone require considerable explanation, and even so we have left out some of the more esoteric experimental facts. It is of historic interest to note that the magnetism of permanent magnets has been known for at least 2,500 years, although the connection of magnetism with moving charges was discovered less than 150 years ago by Oersted, in 1820.

We begin by showing how the very large magnetic effects of ferromagnetism can occur. The ultimate source of magnetic moments in ferromagnetic materials turns out to be (as is primarily the case in paramagnetic materials) the magnetic moments of electron spin. The big difference, however, is that in ferromagnetics there are large interactions *between* spins that cause them to align parallel with each other. Even at room temperatures the torques of interaction are so strong that thermal vibrations cannot destroy the alignment. Thus the very large maximum magnetization is of the same order as would occur in a paramagnetic material if all the dipoles were perfectly aligned along one direction. The source of the interaction between dipoles in ferromagnetics is of quantum-mechanical nature. As a result of the quantum-mechanical torque, the energy of two neighboring dipoles (of two neighboring atoms) is very much less when they are aligned parallel than for any other arrangement. They therefore are highly constrained to take a parallel orientation. Only when a ferromagnetic material is heated to a very high temperature are the thermal motions sufficient to destroy this alignment, thus causing the material to change its behavior to that of a paramagnetic material.

A question that arises immediately is how a ferromagnetic material can ever exist in a nonmagnetized state, in view of the large forces tending to align the dipoles. The solution to this puzzle was suggested long before there was any direct experimental evidence on the subject. The answer given, which has indeed been shown experimentally to be correct, is that there is a strong tendency for the material to break up into *domains* (regions in which all dipoles are aligned), each with a different direction of magnetization, so that the macroscopic effect is to give zero magnetization.

We are still faced, however, with the problem of why the

material chooses to break up into domains, since at their boundaries (the *domain walls*) there will be dipoles that are not parallel. We must account for the forces that pull the dipoles at the boundaries away from the highly desirable parallel arrangement. This phenomenon can be accounted for if we examine the situation from the point of view of stored energy. Figure 9.11 shows schematically

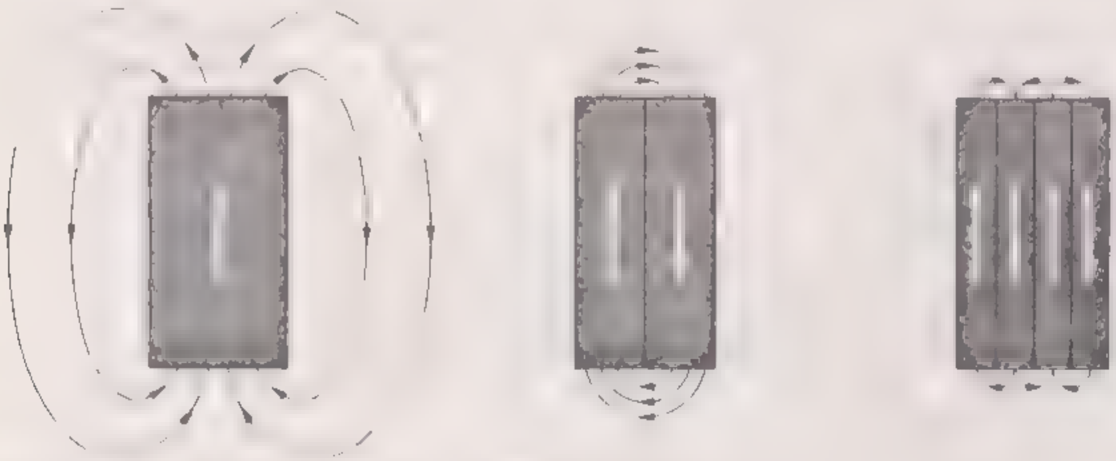


Fig. 9.11 Effect of domains in diminishing the external field.

a number of alternative domain configurations in a piece of material, starting with all the material in a single domain and proceeding to a larger and larger number of domains, which have been chosen in a rather idealized way, to illustrate the idea involved. The orientation of all the dipoles in each domain has been indicated by an arrow. The field outside the material is also indicated in a qualitative manner. A magnetic field outside the material involves stored energy of an amount

$$\frac{1}{2}\frac{B^2}{\mu_0} \qquad \text{joules/unit vol}$$

As the number of domains is increased, the external field produced is smaller and smaller, so that the energy stored in the field is greatly decreased. This decrease must be balanced against the energy stored in making the domain walls, and the equilibrium state is that for which the total energy stored is a minimum. Calculations based on the known work required to disorient adjacent moments indicate that the stable configuration occurs with domains of dimensions of the order of 10^{-5} cm. This agrees with the measured size of domains in a number of ferromagnetics.

We are now in a position to understand the nature of the magnetization curves of ferromagnetic materials. Figure 9.12

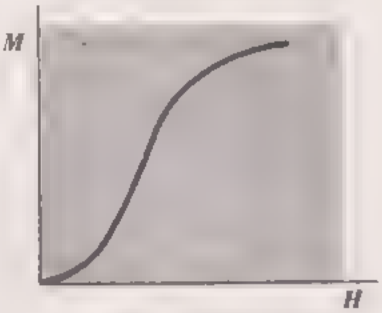


Fig. 9.12 Magnetization curve of a ferromagnetic material.

shows a typical curve of **M** versus **H**. As the magnetizing field is applied, domain walls move so as to favor the growth of domains that happen to have their direction of magnetization more or less along the external field direction. As the field is further increased, forces are great enough to cause the gradual rotation of magnetization direction into exact alignment with the field. There are certain directions of orientation of the dipoles in a domain with respect to the crystal axes that are of lower energy than others, and these directions are taken until the external field forces overcome the internal orienting forces. Finally, when all the dipoles are aligned, **M** has reached a constant value, or is *saturated*. It is often more convenient to plot **B** than **M**, but the difference is often rather small except for the factor μ_0 since $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$, and **M** is usually much greater than the magnetizing force **H**.

A typical experimental curve of **B** versus **H** is plotted in Fig. 9.13, showing the effect of past history on the magnetization of a ferromagnetic sample. We start with the material unmagnetized

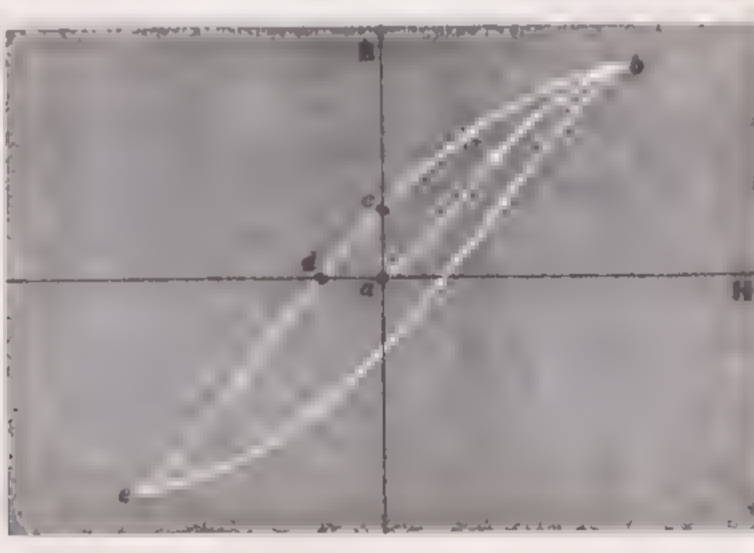


Fig. 9.13 Hysteresis loop.

at a and apply an increasing H (via a current winding around the sample) until the point b is reached. When we decrease **H**, the magnetization **M** or **B** decreases, but along a different path. When **H** has been decreased to zero, we find magnetization remaining (point c). The value of **B** at c is called the *remanence*, which gives the state of permanent magnetization of the sample. Point d gives the *coercive force*, the reverse field necessary to demagnetize the sample. If we continue to point e and reverse the direction of change of **H**, we eventually can trace out a closed loop, called the *hysteresis loop*. The cause of this loop lies in the difficulty of shifting the domain walls. Crystal imperfections cause the walls to tend to stick where they are instead of moving smoothly with the applied force. It is this hysteresis, which is very great in some materials, that allows highly magnetized permanent magnets to exist, a fact of tremendous practical importance. Ferromagnetic materials with large hysteresis are called *hard*, while those having small hysteresis are called *soft*.

Knowledge of the nature of magnetic materials has been greatly increased in recent years by the use of neutrons in the investigation of permanently magnetized materials. The method depends on the fact that neutrons themselves have magnetic moments[1] and that as a result the passage of neutrons through matter is affected by the orientation of the atomic magnetic moments in the matter. Thus neutron diffraction experiments can give information on the relative orientation of the magnetic dipoles of neighboring atoms. Such experiments show the parallel orientation of neighboring dipoles in ferromagnetic materials and the antiparallel orientation of neighbors in *antiferromagnetic* materials. In the latter class of materials, the magnetic moments may be divided into two interpenetrating sublattices in which all moments associated with one lattice are parallel to each other but are antiparallel to the moments on the other sublattice. Such materials do not give rise to external magnetic fields, since the effects of the two sublattices cancel.

Another class of materials, called *ferrimagnetic*, consists of two sublattices with the moments on one lattice contributing larger fields than those on the other, resulting in a net ferromagnetic

[1] It may seem surprising that a neutron, which has a net electric charge of zero, has a magnetic moment. This is explained by the idea that the neutron is a compound structure made up of equal amounts of positive and negative circulating charge, whose contributions to the magnetic moment do not cancel.

effect. One group of such materials is called *ferrite*, which because it happens to be a non-conductor is widely used in high frequency microwave work. The absence of conduction electrons prevents the serious eddy current losses that would otherwise occur in high frequency applications.

9.11 Hysteresis Losses in Magnetic Materials

When a system is carried through a hysteresis loop, a loss of energy results. We examine this problem in the case of the hysteresis loop of a ferromagnetic material. The simplest formulation can be made for a toroidal solenoid sample, which can be magnetized by a solenoidal current density $j^s_{\text{free}} = Ni/L$, where N is the number of turns carrying a current i and L is the length of the solenoid. For this geometry we have seen that

$$H = j^s_{\text{free}} = \frac{Ni}{L} \qquad \text{amp/m} \tag{9.3}$$

We now ask for the amount of work that must be done to build up a current i in the coil, producing a field $\mathbf{B}$ in the material. We first consider the simple case for which $\mathbf{B}$ is linearly proportional to $\mathbf{H}$, as given by $\mathbf{B} = \mu\mathbf{H}$, and then apply the same reasoning to the more complicated situation where $\mathbf{B}$ follows a hysteresis loop. The dissipation of energy in the resistance of the coil is neglected, since we are concerned only with energy in the magnetic matter.

The work necessary to build up a current in the coil results from the induced emf of Faraday induction while the field $\mathbf{B}$ is increasing. Since $\mathcal{E} = -d\Phi/dt = -NA\, dB/dt$, we can write, for the rate of doing work against this emf,

$$P = -\mathcal{E}i = NA\frac{dB}{dt}i \tag{9.32}$$

We can cast this equation in terms of $\mathbf{H}$ by using Eq. (9.3) quoted above, according to which $i = HL/N$. If this relationship is used, the last equation becomes

$$P = LAH\frac{dB}{dt}$$

Since LA is the volume of the sample, we have that the rate at which work goes into magnetic energy, per unit volume of material, is

$$\frac{dW}{dt} = P = H\frac{dB}{dt}$$

or, multiplying by dt, we get a very basic equation,

$$dW = H\,dB \tag{9.33}$$

When we consider the case where $\mathbf{B} = \mu\mathbf{H}$ (Fig. 9.14), we can at once calculate the work per unit volume to build up a magnetic induction field B_0 by means of an applied magnetic intensity H_0. We simply replace H in Eq. (9.33) by B/μ and then integrate from $B = 0$ to $B = B_0$. Thus we find

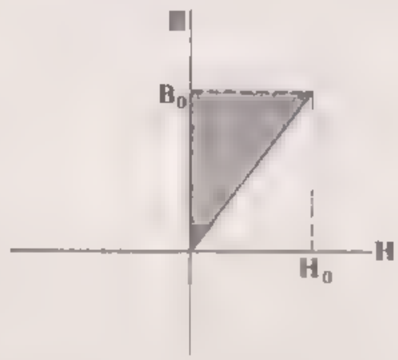


Fig. 9.14 *Evaluation of* $\int_0^{B_0} H\,dB$ *for case where B varies linearly with H.*

$$W = \int dW = \int_0^{B_0} H\,dB = \frac{1}{\mu}\int_0^{B_0} B\,dB = \frac{1}{2}\frac{B_0^2}{\mu} = \frac{1}{2}B_0H_0 \tag{9.34}$$

This is the same result we found in Sec. 9.7, but here we are able to see the relationship between work done and the **BH** curve. $\int_0^B H\,dB$ is simply the shaded area of Fig. 9.14. The factor $\frac{1}{2}$ in the expression is characteristic of all situations where **B** and **H** are linearly related, as was assumed in Sec. 9.7. When the process is reversed, by letting **H** decrease to zero, and if the **BH** curve follows the same path on the way down, an equal amount of energy is fed into the current circuit, since the work integral will have the opposite sign.

On the other hand, if the magnetism follows a **BH** hysteresis loop such as shown in Fig. 9.13 and if we start from a given point and perform a complete cycle back to the original starting point,

the integral of Eq. (9.34) should be taken around a closed loop. In this case the integral has a value just equal to the area of the hysteresis loop. Because of the hysteresis, more energy goes into producing the magnetization than is returned when the magnetization decreases. The lost energy goes into heat.

9.12 Magnetized Bodies

We next discuss the magnetization of bodies in the general case where we are no longer confined to the special toroidal geometry of our previous studies. These discussions will apply equally to paramagnetic, diamagnetic (with appropriate change in the sign of χ_m), and ferromagnetic matter.

In all our previous discussions we have used a toroidal sample shape, where we have shown that inside the magnetic material $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$. The special feature of this shape is that the magnetization vector $\mathbf{M}$ in the material is everywhere parallel to the surface. We shall show that whenever $\mathbf{M}$ has a component perpendicular to the surface, as in the case of all shapes other than toroidal, there result additional effects which influence $\mathbf{H}$ inside and outside the material, and which therefore modify both $\mathbf{M}$ and $\mathbf{B}$. It is still true that the equation $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$ applies, but there are other contributions to $\mathbf{H}$ in addition to the term Ni/L from real currents.

To give a particular example, suppose we have a cylindrical sample of magnetic material magnetized by an externally applied field (Fig. 9.15), with the magnetization vector parallel to the

***Fig. 9.15** Cylindrical magnetic cylinder in an applied field.*

axis. We are going to show that there are contributions to $\mathbf{H}$ both inside and outside the sample due to the effects of the ends. The ends act as sources of $\mathbf{H}$ that must be accounted for in applying the equation $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$ inside the sample, or $\mathbf{B} = \mu_0\mathbf{H}$ outside the sample. This is remindful of the effects of induced polarization surface charges in dielectrics.

We now investigate the effect of the ends of a sample in situations such as shown in Fig. 9.15. We start with Gauss' law for **B**,

$$\int_{CS} \mathbf{B} \cdot d\mathbf{S} = 0 \tag{9.10}$$

As we have seen in Sec. 9.5, this is true in general, whether or not matter is present. The other general statement we have made is that in a magnetic medium, **B** is given by

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) \tag{9.5}$$

When we substitute for **B** in Eq. (9.10) the value of $\mu_0(\mathbf{H} + \mathbf{M})$, we can write

$$\mu_0 \int_{CS} \mathbf{H} \cdot d\mathbf{S} + \mu_0 \int_{CS} \mathbf{M} \cdot d\mathbf{S} = 0$$

or

$$\int_{CS} \mathbf{H} \cdot d\mathbf{S} = -\int_{CS} \mathbf{M} \cdot d\mathbf{S} \tag{9.35}$$

We apply Eq. (9.35) to the right-hand end of the sample of Fig. 9.15. Take a Gaussian pillbox enclosing the end, which has an area A equal to the area of the sample. We find

$$\int_{CS} \mathbf{M} \cdot d\mathbf{S} = -MA \tag{9.36}$$

This follows from the fact that the total number of lines of **M** entering the left-hand side of the pillbox is MA, while the number of lines leaving the pillbox surface outside the sample is zero, since M is zero outside the sample. The negative sign results from the fact that the lines of M are directed *inward* at the Gaussian surface. If this result is substituted in Eq. (9.35), we get

$$\int_{CS} \mathbf{H} \cdot d\mathbf{S} = MA \tag{9.37}$$

This is Gauss' law for **H**. It tells us that, in addition to the closed lines of **H** originating from real currents, there is a net flux of **H** out from any region in which there is a component of **M** perpendicular to a surface. Lines of **H** emerge from the surface (or, in the case of the left-hand end of the sample in Fig. 9.15, where **M** is directed *out* of the Gaussian pillbox in the sample, the net flux of **H** lines is *into* the surface). These extra lines originating or

ending at surfaces modify the field both inside and outside the sample. Furthermore, it can be shown that the lines of **H** from any small region of surface emerge radially outward just as do lines of **E** from a point charge q. We now show this, starting with

$$\oint \mathbf{H}\cdot d\mathbf{l} = i_{\text{free}} \tag{9.13}$$

There are no real currents at the surface, so this becomes $\oint \mathbf{H}\cdot d\mathbf{l} = 0$. The argument is as follows: Since Gauss' law is obeyed by this **H** arising from the discontinuity of **M** at a surface, it follows that lines of **H** coming from this surface effect originate and end only on such surface regions. Thus the concept of lines is valid just as in the electrostatic case for **E**. However, we have not yet shown whether the lines emerge uniformly outward, or whether they may be concentrated in some special direction or directions. Figure 9.16 shows a closed path over which we shall evaluate

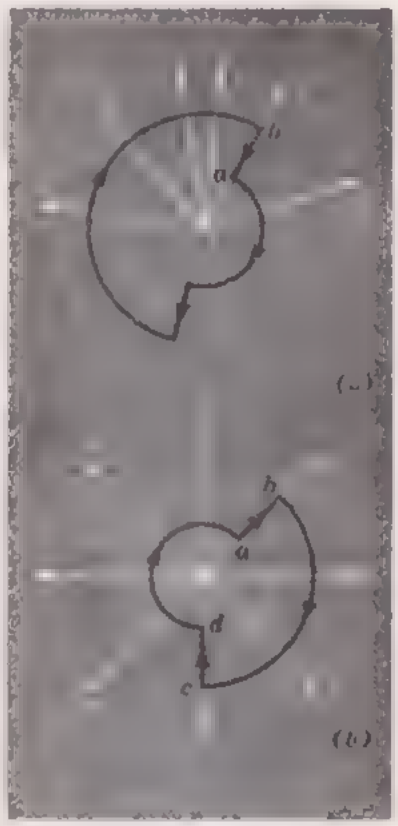


***Fig. 9.16** (a) Hypothetical nonuniform distribution of lines of **H** about a source $\mathbf{M}\cdot d\mathbf{S}$, showing $\oint \mathbf{H}\cdot d\mathbf{l} \neq 0$. (b) True symmetrical distribution of **H** lines about a source, giving $\oint \mathbf{H}\cdot d\mathbf{l} = 0$.*

$\oint \mathbf{H}\cdot d\mathbf{l}$ in the two cases. In case 1 (Fig. 9.16a), where we have assumed a nonuniform distribution of lines of **H**, the line integral is clearly not equal to zero, since the path is parallel to **H** only

between a and b. The path cd is in a region of zero $\mathbf{H}$ and so does not contribute to the line integral. Thus for this case, $\oint \mathbf{H}\cdot d\mathbf{l} \neq 0$. In case 2 (Fig. 9.16*b*), where the lines are uniformly distributed around the point source $\mathbf{M}\cdot d\mathbf{S}$, the path segment ab contributes exactly the same magnitude to the line integral as does the segment cd, and since the two contributions have opposite sign, their net contribution is zero. The remaining segments of the path are perpendicular to $\mathbf{H}$, so they also contribute nothing to the line integral. Thus for uniformly distributed lines of $\mathbf{H}$, $\oint \mathbf{H}\cdot d\mathbf{l} = 0$. This line of reasoning leads directly to the result that Coulomb's law is obeyed. That is, the effects of $\mathbf{H}$ originating from each small end element fall off as $1/r^2$. By analogy with the electrostatic case, we can define the effects of the end discontinuity by

$$-\int_{CS} \mathbf{M}\cdot d\mathbf{S} = q_m \tag{9.38}$$

where we call q_m the *magnetic pole*. Gauss' law for these sources of $\mathbf{H}$ then becomes

$$\int_{CS} \mathbf{H}\cdot d\mathbf{S} = q_m \qquad \text{amp/m} \tag{9.39}$$

In a paramagnetic or ferromagnetic material, $\mathbf{M}$ is in the direction of $\mathbf{H}$. As a result, the extra contribution to $\mathbf{H}$ from the ends is opposed to the $\mathbf{H}$ in the sample and decreases the net value of $\mathbf{H}$ inside. Outside the ends of the sample the lines of $\mathbf{H}$ originating at the sample ends augment the $\mathbf{H}$ due to currents causing the magnetization and thus enhance $\mathbf{B} = \mu_0\mathbf{H}$ outside the sample. In the case of ellipsoidal samples the effects of the ends can be discussed in terms of a *demagnetizing* field, just as we treated the depolarizing field in dielectrics. Although we may discuss other shapes of samples in terms of the demagnetizing factor, results will be only approximate, as was discussed in the case of dielectric materials.[1]

We next discuss the magnetization of two extreme shapes of

[1] When magnetic material is non-uniform, so that the susceptibility is not the same throughout a sample, there can be discontinuities in M within the volume of the sample. Such discontinuities produce an additional term in Eq. (9.37), giving in effect additional sources of lines of H within the volume. We shall not discuss this complication further.

samples. We first take a long, thin rod with its axis parallel to an originally uniform field **B** (Fig. 9.17). If the rod is very thin com-

Fig. 9.17 A long, thin rod of magnetic material with its axis parallel to the original uniform field **B**.

pared with its length, the modification of **H** near the center of the rod due to end effects is very small, since MA is very small when the cross section A is very small. We may reason qualitatively that **H** in the rod is approximately the same as in the absence of the rod. Thus **B** in the rod is $\mathbf{B} = \mu_0(\mathbf{H}_0 + \mathbf{M})$, where $\mathbf{H}_0$ is the value of **H** before inserting the rod. Since $\mathbf{M} = \chi\mathbf{H}_0$, this becomes $\mathbf{B} = \mu_0(1 + \chi)\mathbf{H}_0 = \mu\mathbf{H}_0$. This is the same result we found in the toroidal sample, and is to be expected since here the effects of the sample ends are negligible.

Another line of reasoning leading to the same result is that near the central section of the rod, H_0 outside the rod is essentially unperturbed by the small ends far away and so has its original value. But since H_0 is parallel to the surface in this middle region both inside and outside the rod and since $H_{t1} = H_{t2}$, $H = H_0$ inside the rod. This leads again to $B = \mu H_0$. This result can be described in terms of the demagnetization factor. Thus **H** inside a body is given by

$$\mathbf{H} = \mathbf{H}_0 - L\mathbf{M} \tag{9.40}$$

where L is the demagnetization factor appropriate to any particular sample shape. For the long, thin rod parallel to a uniform field, we have found $L \sim 0$.

We next examine the case of a thin flat plate of magnetic material placed with its surface perpendicular to an originally uniform field (Fig. 9.18). Here the lines of **H** originating at the

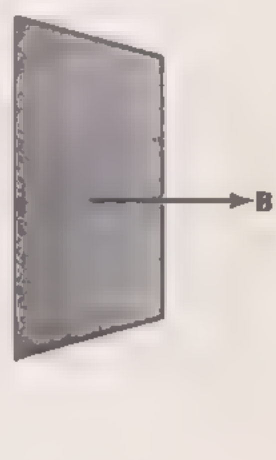


Fig. 9.18 A thin flat plate of magnetic material with its surface perpendicular to the original uniform field.

surface discontinuity of **M** emerge uniform and perpendicular to the surface in both directions, just as lines of **E** from a uniform plane distribution of charge. Since the lines of **H** are parallel, **H** is uniform inside the plate. We can obtain its value inside the plate from

$$\int_{CS} \mathbf{H}\cdot d\mathbf{S} = -\int_{CS} \mathbf{M}\cdot d\mathbf{S} \tag{9.35}$$

using Gauss' law. Thus the contribution from the right-hand surface of the plate, inside the plate, will be $-M/2$ (since half the lines emerge outward). But the left-hand surface will also contribute $-M/2$, so the total **H** inside due to the surfaces is $\mathbf{H} = -\mathbf{M}$. Outside the plate on either side, the contributions from the two surfaces cancel, so **H** outside is unmodified by the plate.

For this case then, inside the plate the demagnetizing field is $-\mathbf{M}$, so the net value of **H** is $\mathbf{H}_0 - \mathbf{M}$, the demagnetization factor L has its maximum value of 1, and

$$\mathbf{B} = \mu_0[(\mathbf{H}_0 - \mathbf{M}) + \mathbf{M}] = \mu_0\mathbf{H}_0 \tag{9.41}$$

The boundary condition for **B**, $B_{n1} = B_{n2}$, is satisfied since **B** is perpendicular to the surface and has the same value inside and outside the plate.

We now examine the case of a magnetic sphere placed in a uniform magnetic field. The resulting **B** field, inside and outside the sphere, is plotted in Fig. 9.19. We are assuming that inside the

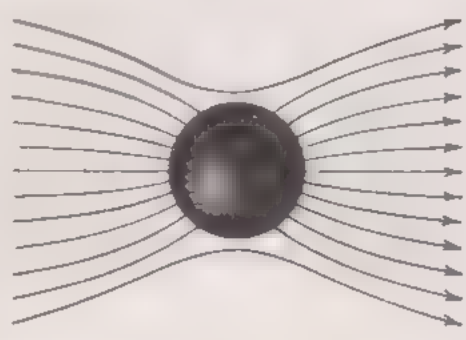


Fig. 9.19 Field perturbation due to magnetized sphere in a uniform field.

sphere $\mu > \mu_0$. The field **B** inside the sphere is increased over its original value and is uniform. This uniformity is to be expected because the sphere is a special case of the general ellipsoid, for which the demagnetizing field is uniform inside the sphere.

In view of the similarity between the dielectric and magnetic equations, it can be argued that the demagnetization factor for

the sphere is the same as the depolarization factor for a dielectric sphere, which is determined in Appendix C. If we use the result found there, $L = \frac{1}{3}$, the field H inside the sphere is

$$H = H_0 - \tfrac{1}{3}M \tag{9.42}$$

where H_0 is the original value in the absence of the sphere. Then we obtain B from

$$B = \mu_0(H + M) = \mu_0[(H_0 - \tfrac{1}{3}M) + M] = \mu_0(H_0 + \tfrac{2}{3}M) \tag{9.43}$$

Since $M = \chi_m H$, we can write $H = H_0 - \frac{1}{3}\chi_m H$ or

$$H = \frac{1}{1 + \frac{1}{3}\chi_m} H_0$$

for the internal H in the sphere. For the general case, this is

$$H = \frac{1}{1 + L\chi_m} H_0 \tag{9.44}$$

Just as in the dielectric case, the perturbation on the *external* field due to the uniformly magnetized field will be exactly the field of a dipole. This applies only outside the sphere, since the field in the sphere is uniform. Although for all these examples of magnetic bodies we have chosen to think in terms of the end effects caused by the discontinuity in M at the body surface, it would be equally satisfactory to replace each magnetized body by the effective magnetic solenoidal surface currents. Thus if a magnetized sphere is replaced by the equivalent surface solenoidal current density $j^s_{mag} = M$, the value of $\mathbf{B}$ inside is $\frac{2}{3}\mu_0\mathbf{M}$ rather than $\mu_0\mathbf{M}$, the value for a toroidal sample. $\mathbf{H}$ inside is thus $\frac{2}{3}\mathbf{M}$ rather than $\mathbf{M}$. Here, without any consideration of the discontinuity of $\mathbf{M}$ at the surface, we find the same result obtained using the ideas of the demagnetizing field of magnetic poles. The two points of view always give identical results.

9.13 Permanent Magnets

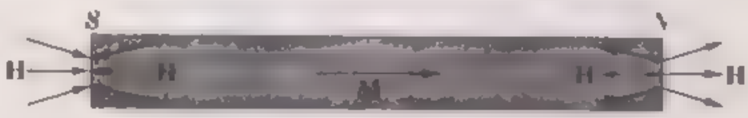
Up to this point we have attributed the magnetization $\mathbf{M}$ to the existence of external fields that caused alignment of the elementary magnetic dipoles in matter. In the case of permanent magnets we must consider that the alignment of dipoles results from internal forces within the matter. We then inquire as to the resulting $\mathbf{B}$

and **H** fields that this self-magnetization produces, both inside and outside the matter.

We again use the equation $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$ for calculating the field inside the material, but since $\mathbf{H}_0$, the applied field, is zero, **H** arises only from end effects from the permanent magnetization **M**, as given by the equation $\int_{CS} \mathbf{H} \cdot d\mathbf{S} = -\int_{CS} \mathbf{M} \cdot d\mathbf{S}$. In the simple case of ellipsoidal shapes, as we have seen, $H = H_0 - LM$, but since H_0 is zero, this becomes $H = -LM$ inside the sample. Thus **H** inside the sample is opposed to **M** (and to **B**) and in fact is just the demagnetizing field we considered in the last section. Only for ellipsoidal samples in which **M** is uniform will **H** be uniform. Outside the sample **H** will not be uniform, but it can be calculated on the basis of effective poles as discussed below.

For nonellipsoidal shapes, the field will not be uniform inside or outside the magnet. However, we can calculate the values of **B** and **H** by using Coulomb's law for the effective poles. Consider the case of a bar magnet as shown in Fig. 9.20, in which we assume

***Fig. 9.20** Permanent bar magnet showing sources and sinks of lines of **H** at ends of magnet.*



M is uniform. We have shown the lines of **H** originating and ending at the discontinuities in **M** at the ends of the magnet. The strength of these poles, we have seen, is given by

$$q_m = -\int_{CS} \mathbf{M} \cdot d\mathbf{S} = MA \tag{9.38}$$

where A is the cross section of the magnet. The end of the magnet from which lines of **H** emerge is called the north pole, and the end on which lines of **H** converge is called the south pole. When we examine the field at a point far enough away from the end of the magnet so that the size of the end region is small compared with the distance from the region to the point in question, the pole can be taken as a point. Applying Gauss' law for **H**, $\int_{CS} \mathbf{H} \cdot d\mathbf{S} = q_m$, over a Gaussian sphere of radius r gives us

$$H = \frac{1}{4\pi} \frac{q_m}{r^2} \qquad \text{or} \qquad B = \frac{\mu_0}{4\pi} \frac{q_m}{r^2} \tag{9.45}$$

This gives **B** anywhere outside the rod due to the pole q_m. Since there are always both a north and a south pole associated with a magnetized body, the net field at any point is the vector sum of contributions from both poles. This is exactly the same situation we had in electrostatics when we calculated the electric field from an electric dipole. Since lines of **H** emerge from a north pole, q_m is positive for a north pole and negative for a south pole. If the point at which we are examining the field is far away from the magnet relative to its length L, the equations for **B** or **H** reduce to the usual dipole equation (6.24), where the magnetic dipole is $p_m = q_m L$.

The torque on a permanent bar magnet placed in a uniform field, using the magnetic-dipole equation, is

$$\tau = p_m B \sin \theta \tag{6.23}$$

Furthermore, we can obtain the expression for the force on a magnetic pole. We modify the torque equation by replacing p_m by $q_m L$, to obtain

$$\tau = q_m L B \sin \theta \tag{9.46}$$

This is exactly analogous to the case of an electric dipole in an electric field, where the force is $F = qE$. By analogy, the force on a magnetic pole must be

$$F = q_m B \tag{9.47}$$

For a north pole $(+)$, the force is parallel to **B**; for a south pole $(-)$, the force is antiparallel to **B**.

We have now developed the field and force equations for permanent bar magnets. We can calculate the field anywhere in space due to an assembly of permanent magnets through Eq. (9.45) and can calculate the force on each pole in a given field through Eq. (9.47). This treatment is possible only if the magnetization of the magnets is constant and not affected by external fields. For many situations this is a good approximation. In principle, all the results we have obtained by using the concept of magnetic poles could have been obtained by replacing each magnet by an equivalent solenoidal surface current.

A word must be said about the nomenclature of north and south magnetic poles and the connection between this and the mag-

netic field of the earth. A more correct name for the north pole of a magnet as we have defined it would be the *north-seeking* pole. The south pole of a magnet is more correctly the *south-seeking* pole. Since opposite poles attract, this means that the magnetic pole near the north of the earth is a south magnetic pole. The magnetic field of the earth is approximately described in terms of a magnetic dipole whose axis is roughly parallel to the axis of rotation of the earth. The horizontal component of the earth's field points generally in the south-north direction. The principal source of the earth's field is probably the convection currents of molten conducting matter in the central volume of the earth. This motion gives rise to an electric current within the earth and a consequent magnetic field. It is unlikely that ferromagnetic materials play a dominant role in the earth's magnetism, since it is expected that they lose their very large susceptibility and become paramagnetic at the high temperatures in the interior of the earth.

9.14 Examples

a The field in and around a permanently magnetized rod In Sec. 9.12 we found that the demagnetization factor L for this shape is about zero. Therefore, inside the rod $H = H_0 - LM$ is approximately zero, since the externally applied field is zero. Whatever H does exist in the rod is opposed to M. Outside the rod, the field can be calculated by the use of Eq. (9.45), $B = \frac{\mu_0}{4\pi}\frac{q_m}{r^2}$, applied to both ends, where $q_m = MA$. The resulting field is shown in Fig. 9.21

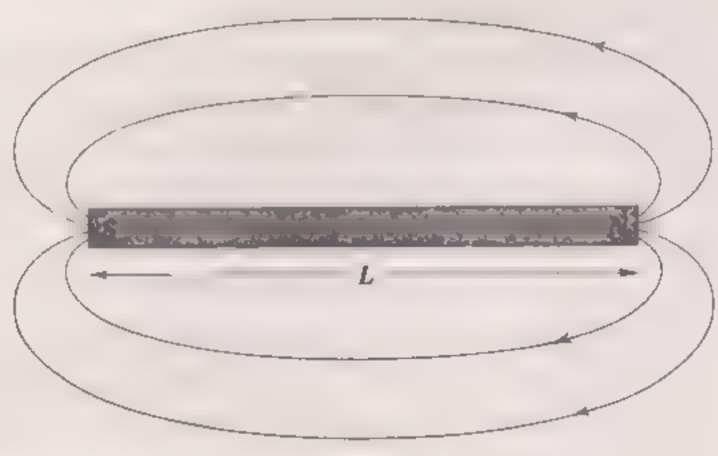


Fig. 9.21 ***Bar magnet. External field can be described by poles of strength q_m separated by a distance L.***

b The field in and around a thin plate permanently magnetized perpendicular to its plane For this magnetization, we found in

Sec. 9.12 that the demagnetization factor is 1. The equation for **B** inside the plate was

$$\mathbf{B} = \mu_0[(\mathbf{H}_0 - \mathbf{M}) + \mathbf{M}] = \mu_0\mathbf{H}_0 \tag{9.41}$$

so in this case, where the applied field $\mathbf{H}_0 = 0$, $\mathbf{B} = 0$ inside the plate and $\mathbf{H} = -\mathbf{M}$. Thus **H** inside the plate is opposed to **M** and equal to it in magnitude. Outside the plate, since $B_{n1} = B_{n2}$, **B** is also zero, as is **H**. We neglect the perturbing effects of the edges of the plate.

c *The field in and around a uniformly magnetized sphere* We found in Sec. 9.12 that the field inside a sphere of magnetization **M** in an external field is

$$\mathbf{B} = \mu_0(\mathbf{H}_0 + \tfrac{2}{3}\mathbf{M}) \tag{9.43}$$

In the absence of an external field then, $\mathbf{B} = \frac{2}{3}\mu_0\mathbf{M}$. This result followed from the fact that the demagnetizing factor for a sphere, as shown in Appendix C for the dielectric case, is $L = \frac{1}{3}$. The field inside the sphere is then $\mathbf{H} = -\frac{1}{3}\mathbf{M}$. **H** is uniform and opposed to the permanent magnetization **M**. Outside the sphere we have argued that the shape of the field is that of a dipole. We can determine the magnitude of the dipole moment by matching boundary conditions at the surface of the sphere at the point on the sphere along its axis of magnetization. This is the point P in Fig. 9.22. At this point, **B** is normal to the surface, so **B** inside

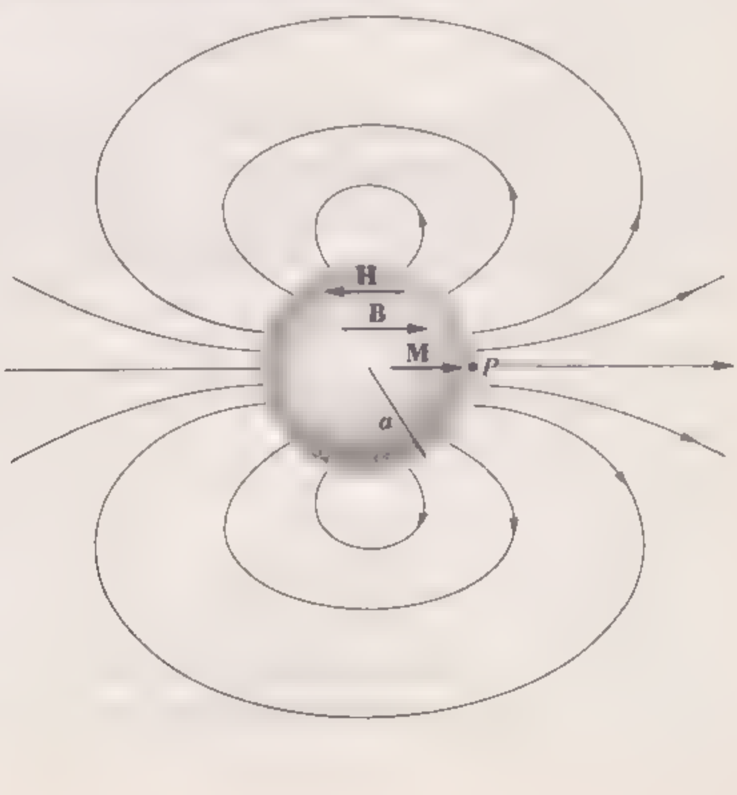


***Fig. 9.22** Magnetic field in region around a uniformly magnetized sphere; M, B, and H are uniform inside the sphere. Note that H is opposite to B inside the sphere.*

and just outside the surface must have the same value (since $B_{n1} = B_{n2}$). The value of **B** inside the surface at P is $\frac{2}{3}\mu_0\mathbf{M}$. Using the dipole formula outside the sphere and setting $\cos\theta = 1$, we have from Eq. (6.24)

$$B_r = \frac{\mu_0}{4\pi}\frac{2p_m}{a^3} \tag{9.48}$$

where a is the radius of the sphere. Comparing these two expressions, we can solve for p_m to obtain

$$p_m = \tfrac{4}{3}\pi a^3 M \tag{9.49}$$

The value of **B** anywhere outside the sphere may now be obtained by taking the vector sum of the two perpendicular components of **B**, $\mathbf{B}_r$, and $\mathbf{B}_\theta$, obtained by substitution of this value of p_m in the dipole equations (6.24). The dipole moment of the sphere is just its magnetization times its volume.

d Calculation of the field inside a permanently magnetized sphere by means of the effective solenoidal surface current. We show this calculation as an illustration of the fact that the effect of magnetized bodies can be simulated by an effective solenoidal current density at the surface of the body. We found in Sec. 9.2 that the magnetic effects of a uniformly magnetized body can be simulated by the Amperian surface current density j^s_{mag}, where

$$j^s_{mag} = M \tag{9.2}$$

Figure 9.23 shows a few of the turns of wire wound around the sphere in which a current i produces the field equivalent to that of the Amperian surface current. The turns are wound uniformly along the z axis in order to give a uniform solenoidal current density. We

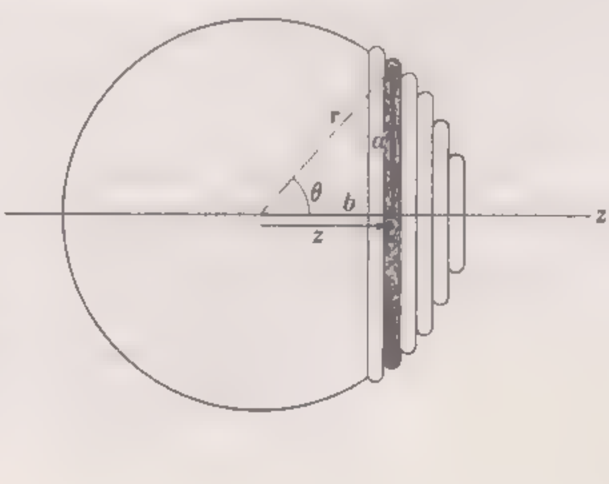


Fig. 9.23 Uniform solenoidal current winding on surface of a sphere

now calculate the field due to this current, at the center of the sphere. We calculate the contribution from one turn and then integrate over all the turns on the sphere.

The contribution from the one turn we have chosen is given by

$$dB = \frac{\mu_0 i}{2} \frac{a^2}{(a^2 + b^2)^{3/2}} = \frac{\mu_0 i}{2} \frac{a^2}{r^3} \tag{6.8}$$

If z is the distance along the z axis from the center of the sphere to a given turn, $z = r \cos \theta$ and $dz = r \sin \theta \, d\theta$. If dz is the thickness of one turn, then $i = j\,dz$ Also, the radius of the turn is $a = r \sin \theta$. Substitution of these quantities into Eq. (6.8) gives

$$dB = \frac{\mu_0 j}{2} \frac{r^3 \sin^3 \theta \, d\theta}{r^3}$$

or

$$B = \frac{\mu_0 j}{2} \int_0^\pi \sin^3 \theta \, d\theta = \frac{\mu_0 j}{2} \left[-\frac{1}{3} \cos \theta \, (\sin^2 \theta + 2) \right]_0^\pi$$

$$= \tfrac{2}{3} \mu_0 j$$

Since the current density j is to simulate j^s_{mag}, we set it equal to M. Then we have

$$B = \tfrac{2}{3} \mu_0 M \tag{9.50}$$

which is the value found in the last problem, using the method of magnetic poles. Note that our proof is only for the point at the center of the sphere. We could show, however, that B is indeed uniform throughout the sphere.

9.15 Magnetic Circuits

The general problem of magnetic bodies in external fields is extremely difficult. We are involved in the simultaneous solution of

$$\oint \mathbf{H} \cdot d\mathbf{l} = i_{free} \tag{9.13}$$

$$\int_{CS} \mathbf{B} \cdot d\mathbf{S} = 0 \tag{9.10}$$

$$\mathbf{B} = \mu \mathbf{H} \tag{9.9}$$

and the boundary conditions on **B** and **H** [Eqs. (9.16) and (9.18)]. However, there is one kind of situation involving ferromagnetic materials that is both important practically and very easy to solve approximately. An example is shown in Fig. 9.24. This is an electro-

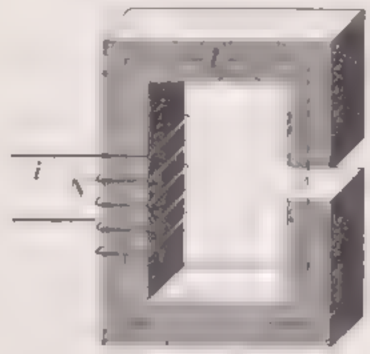


Fig. 9.24 Electromagnet, illustrating a magnetic circuit.

magnet. The problem is to determine the value of **B** anywhere in the path through the magnetic circuit shown. In particular, we might wish to know the value of **B** in the air gap. We shall assume we know the magnetizing current i through the N turns of the coil, the cross-section area A of all sections, and the value of μ for all parts. We must further assume that the lines of **B** are parallel to and confined to the circuit of matter. This is approximately true if the material used has a large μ (as is the case for ferromagnetic materials).

We first use Eq. (9.13) and take the line integral around the circuit,

$$\oint \mathbf{H}\cdot d\mathbf{l} = H_1l_1 + H_2l_2 + H_3l_3 + \cdots = Ni \tag{9.51}$$

We have here divided the line integral into sections, each taken over a length having a constant value of μ and a constant cross-section area. The $\cos\theta$ term of the scalar product goes to 1 because **H** is everywhere parallel to the path taken. Equation (9.10) tells us that the lines of **B** are continuous around the circuit. Therefore the flux Φ passing through any cross section of the circuit is the same. We can relate the value of H in any section to the flux, using Eq. (9.9) to give

$$\Phi = BA = \mu HA \tag{9.52}$$

where the μ and A must be taken for that section. We thus find

$$H_1 = \frac{\Phi}{\mu_1 A_1} \qquad H_2 = \frac{\Phi}{\mu_2 A_2} \cdots$$

Substitution in Eq. (9.51) then gives

$$Ni = \Phi\left(\frac{l_1}{\mu_1 A_1} + \frac{l_2}{\mu_2 A_2} + \cdots\right) \tag{9.53}$$

This equation can be solved for Φ, after which $B = \Phi/A$ can be found for any section required, in particular for the air gap in our example of an electromagnet (for the air gap we use $\mu = \mu_0$). We may think of Eq. (9.53) as a circuit in analogy to a series electric circuit, $V = i(R_1 + R_2 + \cdots)$. Just as current is continuous, in the magnetic case the flux Φ is continuous. The terms $l/\mu A$ are of the same form as resistance and combine in series and parallel in the same way. They are often called the *magnetic reluctance* $\mathcal{R}$. The driving force for the magnetism is Ni called the *magnetomotive force* (mmf). The equation can be written as

$$\Phi = \frac{\text{mmf}}{\mathcal{R}} \tag{9.54}$$

It is easy to see from the foregoing ideas why the magnetic field of an electromagnet can be increased by tapering the poles of the magnet. The total reluctance of the magnetic circuit is almost unchanged by tapering the poles, so the flux is almost unaltered. However, the effective cross-section area of the air gap is greatly reduced so that $B = \Phi/A$ will be much increased. If the tapering is too sharp, the lines of B will not be pulled in, and improvement may be only slight.

PROBLEMS

9.1 A toroidal sample of magnetic material of susceptibility $\chi_m = 2 \times 10^{-2}$ is wound with 1,000 turns of wire carrying a current of 2 amp. The toroid is 15 cm long.

a Find the solenoidal current density j^s_{free}.

b Determine the magnetic field intensity H produced by the current.

c Calculate μ, the magnetic permeability of the material.

d Calculate the induced magnetization M in the material.

e Calculate the magnetic induction field B resulting from the current and the magnetization of the material.

9.2 A rod of paramagnetic material of uniform cross section A is placed in a nonuniform magnetic field between the poles of a magnet, as shown in Fig. P9.2. Calculate the vertical force on the rod as a result of the nonuniform field. Proceed by calculating the effect on the stored magnetic energy of a small displacement dx of the rod in the direction of its axis. To simplify the problem, consider that instead of moving the entire rod, a thin slice of thickness dx is removed from the bottom of the rod and added to the top. If the field at the bottom

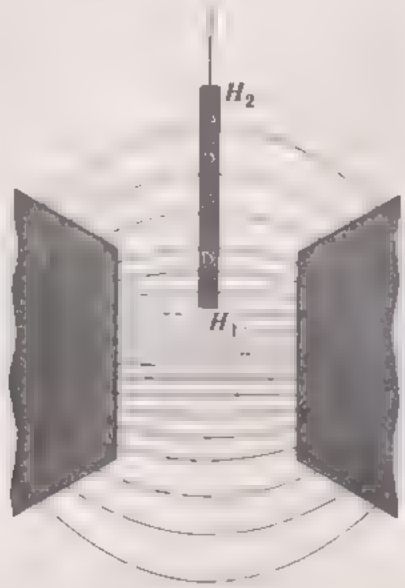


Fig. P9.2

of the rod is H_1 and at the top is H_2, show that the *change* of magnetic energy is given by (see Sec. 9.7)

$$dW = \tfrac{1}{2}A(\mu - \mu_0)(H_1^2 - H_2^2)\,dx$$

Using this result, show that the force on the rod is

$$F = \tfrac{1}{2}A\mu_0\chi(H_1^2 - H_2^2) \qquad \text{newtons}$$

This is a method often used for measuring the static susceptibility of paramagnetic substances. Note that the force is in the opposite direction if the material is diamagnetic.

9.3 An iron ring of radius 5 cm has a cross section of 2 cm^2. Its permeability is 1,000 μ_0 (assumed constant). It is wound with 1,500 turns carrying 5 amp. Calculate the following:

a The self-inductance of the coil
b The magnetic field intensity
c The solenoidal current density of the coil
d The magnetization of the iron
e The induced Amperian surface current density
f The stored magnetic field energy

9.4 Two uniformly magnetized rods have length L and cross-section areas A. They have a magnetization M amp/m. Find their pole strengths q_m. One magnet is suspended above the other by the repulsion of like poles. The mass of each magnet is m kg. Find the equation whose solution will give equilibrium spacing x between the rods.

9.5 Two similar permanent magnets A and B, having magnetic moments p_m, are arranged as shown in Fig. P9.5. They are separated by a distance large compared with their lengths. A compass needle is placed a distance x_1 from one magnet and x_2 from the other. The compass needle takes up a position θ as shown in the figure. Find the ratio x_1/x_2.

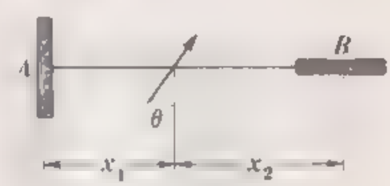


Fig. P9.5

9.6 The electromagnet shown in Fig. P9.6 is made of iron with a magnetic permeability $\mu = 800\ \mu_0$. The cross section of the iron is 10×10 cm, and the length of the path around the iron and across the 2-cm gap is 200 cm. How many ampere-turns will be required in the windings to give a 5,000-gauss (0.5 weber/m^2) field in the gap, assuming no bulging of magnetic field lines out of the gap? What is B inside the iron? What is H in the gap? What is H in the iron? What is the magnetization M in the iron?

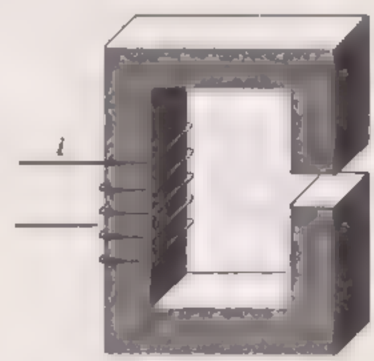
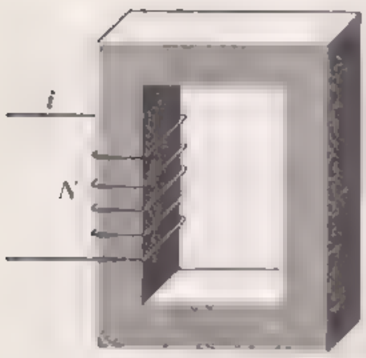


Fig. P9.6

9.7 Metallic iron contains approximately 10^{29} atoms per cubic meter. The magnetic moment of each iron atom is 1.8×10^{-23} amp-m^2. If there were no internal ferromagnetic forces tending to align the dipoles (i.e., if iron were paramagnetic), what would be the susceptibility of iron at 300°K? What would be the dipole moment of an iron bar (if it were paramagnetic) of dimensions 10 cm long and 1 cm^2 cross section in a field of 1,000 gauss (0.1 weber/m^2)? If all the dipoles were aligned in one domain as a result of ferromagnetic interactions, what would be the magnetization M in the bar? What would be the magnetic moment of the bar? What would be the torque on this ferromagnetic bar in a field of 100 gauss perpendicular to the axis of the

bar? What would be the magnetic pole strength of the ferromagnetic bar?

9.8 An inductance is formed of 100 turns of wire wrapped around a closed iron loop 20 cm in length and of 1×1 cm cross section, as shown in Fig. P9.8. A 60-cps alternating current is passed through the



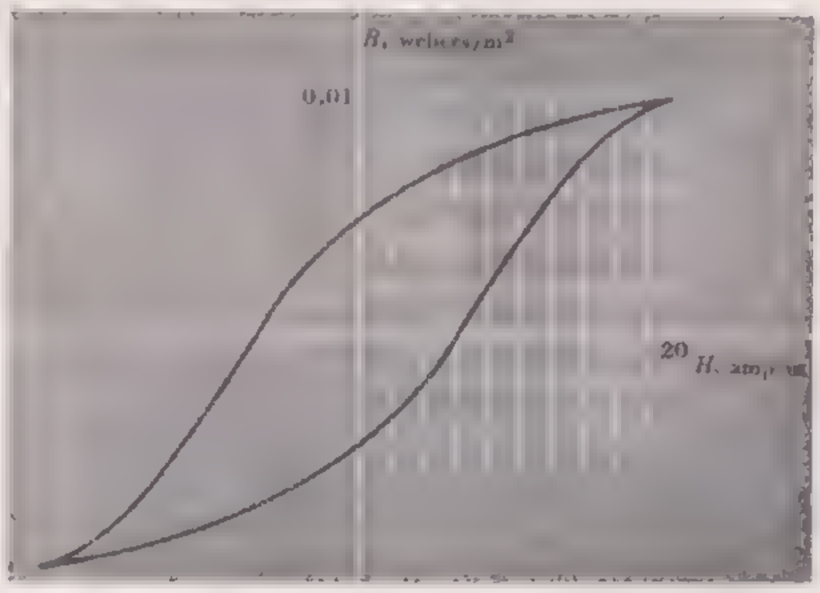


Fig. P9.8

coil. The iron goes through the hysteresis loop once each cycle. Find the approximate power loss due to hysteresis. Use the expression for $\oint \mathbf{H} \cdot d\mathbf{l}$ to obtain the peak current in the coil from information given on the hysteresis plot. What is the self-inductance of the inductor under the conditions of the problem? If twice the current were passed through the coil, how would this affect the inductance (qualitative)?

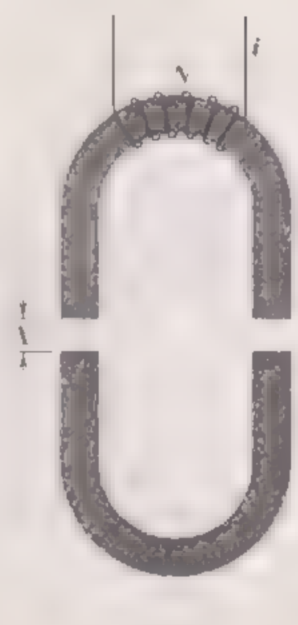


Fig. P9.9

9.9 An electromagnet is constructed as shown in Fig. P9.9. The total length of iron in the two parts is L, the cross-section area is A, and the permeability of the iron is μ. A current i through N turns activates the magnet. The two halves of the magnet are separated by a distance X, where $X \ll A$. Find the force of attraction between the two halves of the magnet (see Sec. 9.7).

9.10 A variable inductance is made by winding a coil on an iron core 30 cm long, which has an adjustable air gap. With no air gap, the self-inductance is 2 henrys. For what air gap is the inductance reduced to 1 henry? The permeability of the iron is 1,000 μ_0.

TEN

Alternating-current Circuits

10.1 Introduction

We turn next to a consideration of circuits. This involves no new electric or magnetic principles but depends on the characteristics of the three passive circuit elements R, C, and L, which have already been investigated. The a-c circuit is of major importance in applied electricity, though here we shall limit our discussion to elementary circuits and the study of some simple methods for their analysis. Our objective is to be able to describe currents and voltages in a circuit such as shown in Fig. 10.1 when it is connected to a sinusoidal source of voltage. We shall develop the ideas of phase and amplitude of sinusoidal functions, which are used to describe

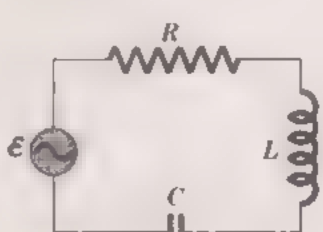


Fig. 10.1 Series a-c circuit.

the behavior of the circuit and its individual components. The analysis of a-c circuits involves the setting up and solving of certain differential equations. In order to gain insight into the behavior of such circuits, we shall examine the problem from several points of view.

In addition to developing the ideas necessary for discussing current-voltage relationships in a-c circuits, we give brief mention to a-c filter circuits, discuss power dissipation in a-c circuits, and list a few instruments used for measuring a-c current, voltage, and power. We conclude this chapter with short discussions of the transformer, generators of electric power, and motors for converting electric energy into mechanical energy.

10.2 Sinusoidal Voltage

In Chap. 8 we saw that according to Faraday's law of induction, an emf $\mathcal{E}$ is induced in a coil of N turns, area A, rotating with an angular velocity ω in a uniform field $\mathbf{B}$:

$$\mathcal{E} = NAB\omega \sin \omega t \tag{8.20}$$

Such a coil rotating at a constant frequency amounts to an a-c generator, such as might be used as a source of voltage in the circuit of Fig. 10.1. Before proceeding further, we discuss various models and mathematical ways of describing such a voltage. We rewrite Eq. (8.20) as $\mathcal{E} = \mathcal{E}_0 \sin \omega t$, or since the emf develops a voltage v across the circuit, we may write

$$v = V_0 \sin \omega t \tag{10.1}$$

for convenience and plot its meaning in Fig. 10.2. V_0 is called the

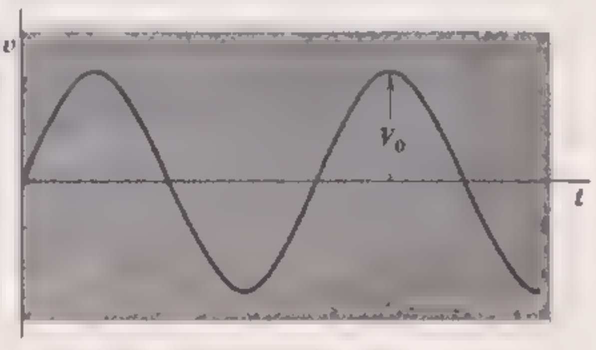


Fig. 10.2 Time variation of a sinusoidal voltage of amplitude V_0.

amplitude, ω the *angular frequency* ($\omega = 2\pi f$, where f is the *frequency*), and the *argument* ωt of the sine gives the *phase* of the voltage.

We may understand this simple terminology by constructing a generating circle as shown in Fig. 10.3. Imagine the wheel shown

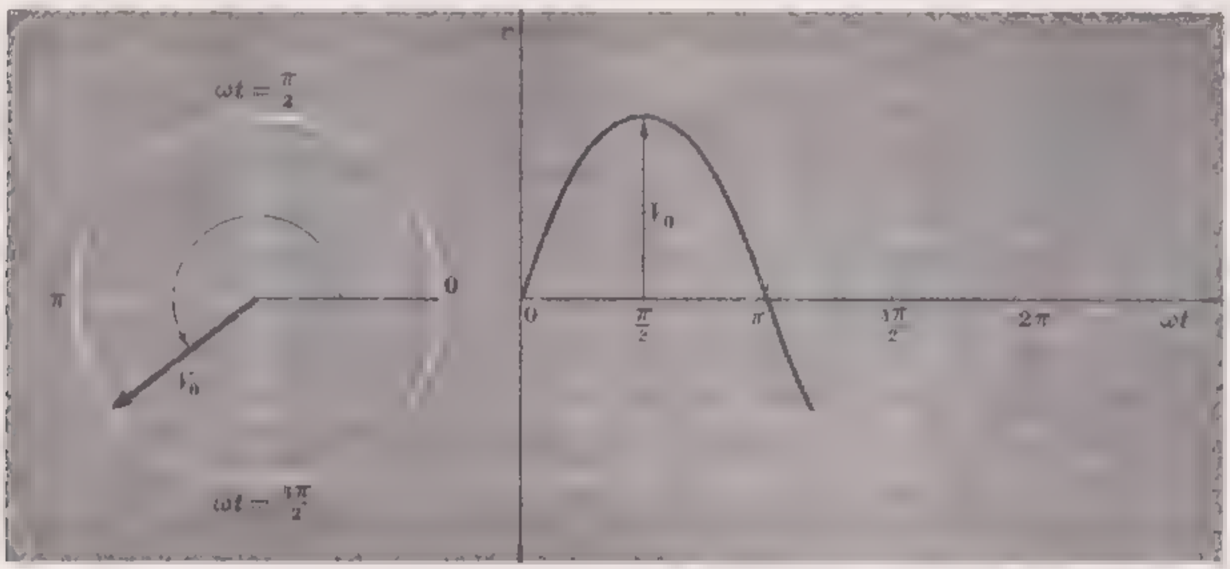


Fig. 10.3 Generating circle forms sine function.

as rotating at a frequency f rps or $\omega = 2\pi f$ radians/sec, in a counterclockwise direction. A line is painted on the wheel, and at the time $t = 0$ this line is in the position marked $\omega t = 0$. The point of projection on the y axis moves sinusoidally up and down, and if we move the projected point along the t axis as shown, the sine function is traced. The terms *amplitude*, *phase*, *frequency*, and *angular frequency* are obvious here.

We could have arrived at the same result by using the equation

$$v = V_0 \cos \omega t \tag{10.2}$$

The only difference is in the starting point of the function at $t = 0$, which is zero for the sine function and the maximum value V_0 for the cosine function. Clearly, either equation can be used here.

In the following discussions, emphasis is given to the behavior of circuits when a *sinusoidal* voltage is applied. Although there are many more complicated kinds of voltage variation, there are two reasons for concentrating on this simple form. One is that a large fraction of the practical applications of alternating current involve simple sinusoidal voltage and current. The other is that the best approach to the solution of circuit behavior for more complicated time variations is to analyze the time variation in terms of the superposed effects of appropriate sine and cosine

functions. Thus the basic problem in all cases is to understand the behavior of circuits for sinusoidal applied voltages. The analysis of the behavior of circuits in terms of superposed sine and cosine functions depends on the *linear* nature of the basic circuit elements. For example, we shall find that when the a-c voltage applied to a resistance, capacitance, or inductance is doubled, the a-c current is also doubled.

10.3 Current-Voltage Relations

Before solving for the behavior of a circuit such as in Fig. 10.1, we summarize what we know about the relationship between voltage and current when a sinusoidal voltage is applied to the three circuit elements separately.

Resistance The equation for the circuit shown in Fig. 10.4 is

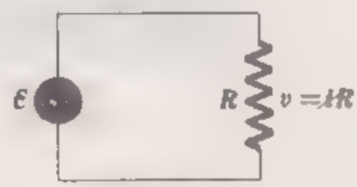


Fig. 10.4 Circuit with emf and resistance.

$\mathcal{E} = v = iR$. Since R is a constant, if either v or i is sinusoidal, the other is also and has the same phase. For instance, we may take as the sinusoidal voltage $v = V_0 \cos \omega t$. Substitution in the circuit equation gives for the current

$$i = \frac{V_0}{R} \cos \omega t \tag{10.3}$$

The amplitude of this sinusoidal current is thus V_0/R. If we call this current amplitude I_0, we have

$$I_0 = \frac{V_0}{R} \tag{10.4}$$

Since both i and v are proportional to $\cos \omega t$, the current through the resistor is in phase with the voltage across it. According to Eqs. (10.3) and (10.4), Ohm's law is valid not only for the instantaneous relationship between i and v but also for the sinusoidal amplitudes I_0 and V_0.

We can show the results above graphically as in Fig. 10.5. Here the two lines V_0 and I_0 on the rotating wheel give the amplitudes and relative phase of the two sinusoidal quantities v and i. When we also fix the angular frequency ω of the wheel, we know

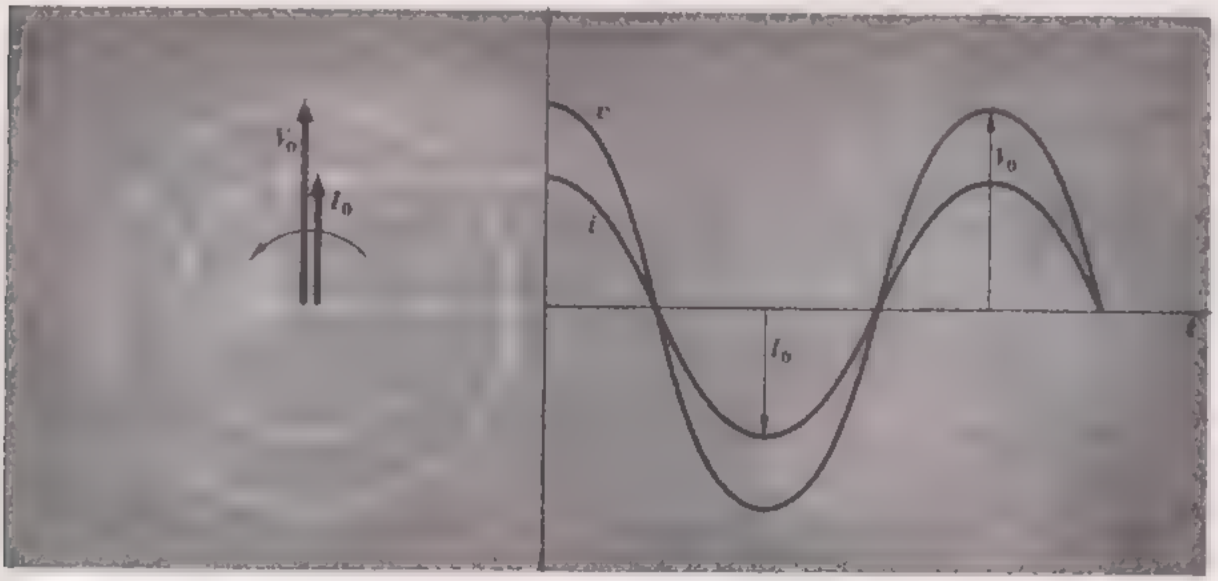


***Fig. 10.5** Current-voltage relationship in a resistance. Current and voltage are in phase.*

all there is to know about v and i. We call these lines V_0 and I_0 on the wheels the *generating vectors* for voltage and current.

Inductance We show in Fig. 10.6*a* an inductance connected to an

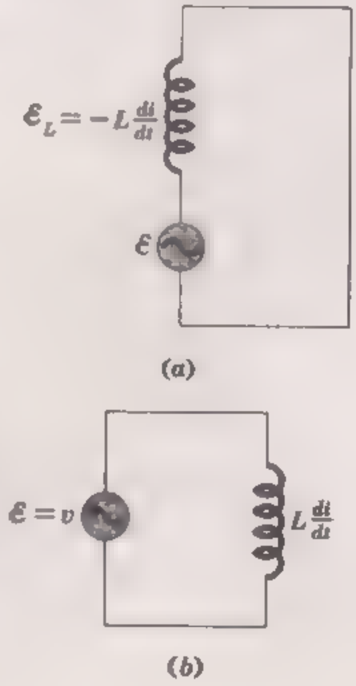


***Fig. 10.6** A pure inductance connected to an alternating emf. (a) The circuit equation gives $\mathcal{E} + \mathcal{E}_L = 0$. (b) For an applied voltage $\mathcal{E} = v$, the potential drop across the inductance is $v = L\,di/dt$.*

alternating emf $\mathcal{E}$. The emf induced across the inductance is given by

$$\mathcal{E}_L = -L\frac{di}{dt} \tag{8.17}$$

The circuit equation is then

$$\mathcal{E} + \mathcal{E}_L = 0$$

This gives

$$\mathcal{E} - L\frac{di}{dt} = 0 \qquad \text{or} \qquad \mathcal{E} = L\frac{di}{dt}$$

We may redraw the circuit as shown in Fig. 10.6*b*. Then, since the alternating emf produces an alternating voltage v across the terminals of the generator, the last equation can be written as

$$v = L\frac{di}{dt} \tag{10.5}$$

Thus an applied voltage v produces a changing current in the inductance. The magnitude of di/dt is given by Eq. (10.5).

If we take $v = V_0 \cos \omega t$, solving for di/dt gives

$$\frac{di}{dt} = \frac{V_0}{L}\cos \omega t$$

Upon integration, we obtain the expression for current,

$$i = \frac{V_0}{\omega L}\int \cos \omega t\, \omega\, dt = \frac{V_0}{\omega L}\sin \omega t \tag{10.6}$$

When, as before, we let I_0 be the current amplitude, we have

$$I_0 = \frac{V_0}{\omega L} \tag{10.7}$$

The quantity ωL plays the role of R in Eq. (10.4). This is called the *inductive reactance*. Equation (10.7) has the same form as Ohm's law, but the current and voltage do not have the same phase. If v

varies as $\cos \omega t$, i varies as $\sin \omega t$. This result is plotted in Fig. 10.7. From the graphical model we see that the phase of the current *lags*

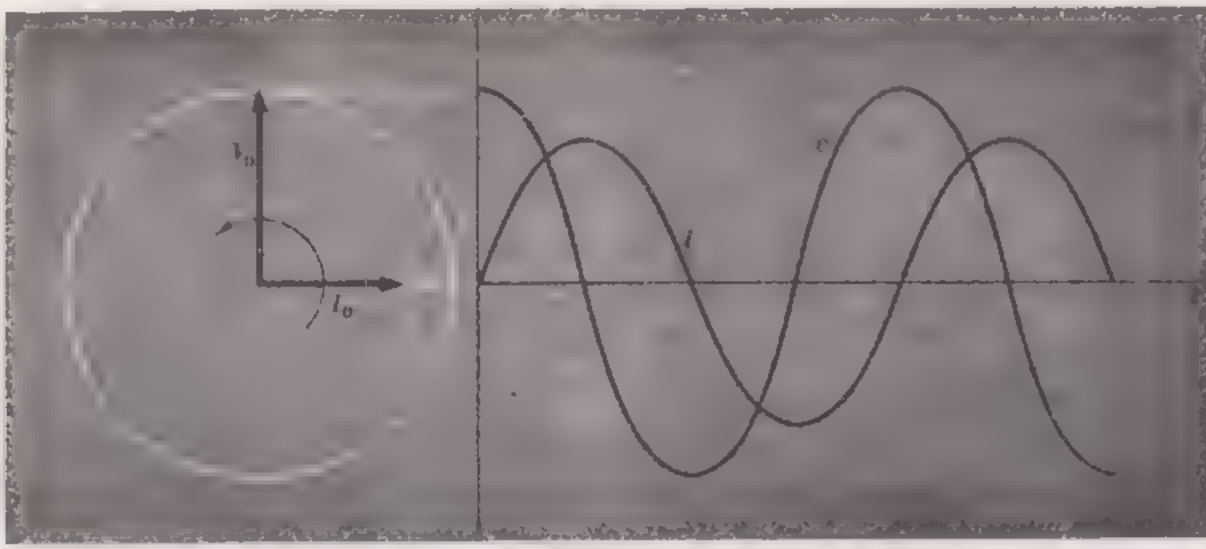


***Fig. 10.7** Current-voltage relationship in an inductance. Current lags the voltage by a phase angle of 90°.*

the voltage source by a phase angle of 90°. We can also get this from the trigonometrical transformation,

$$i = \frac{V_0}{\omega L} \sin \omega t = \frac{V_0}{\omega L} \cos\left(\omega t - \frac{\pi}{2}\right)$$

Capacitance The circuit equation for a capacitance (Fig. 10.8)

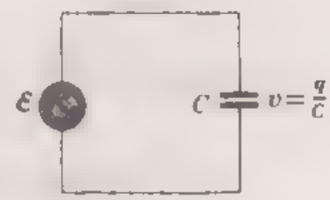


***Fig. 10.8** Circuit with emf and capacitance.*

is $\mathcal{E} = v = q/C$. Using $i = dq/dt$, we convert this by differentiation to

$$\frac{dv}{dt} = \frac{1}{C} i \tag{10.8}$$

Again choosing $v = V_0 \cos \omega t$, we find $dv/dt = -\omega V_0 \sin \omega t$, so

$$i = -\omega C V_0 \sin \omega t \tag{10.9}$$

A plot of this is given in Fig. 10.9. The amplitude of the current is

$$I_0 = \omega C V_0 \tag{10.10}$$

Here the term that replaces R in Ohm's law is $1/\omega C$. This quantity is the *capacitative reactance*. In this case the phase of the current is

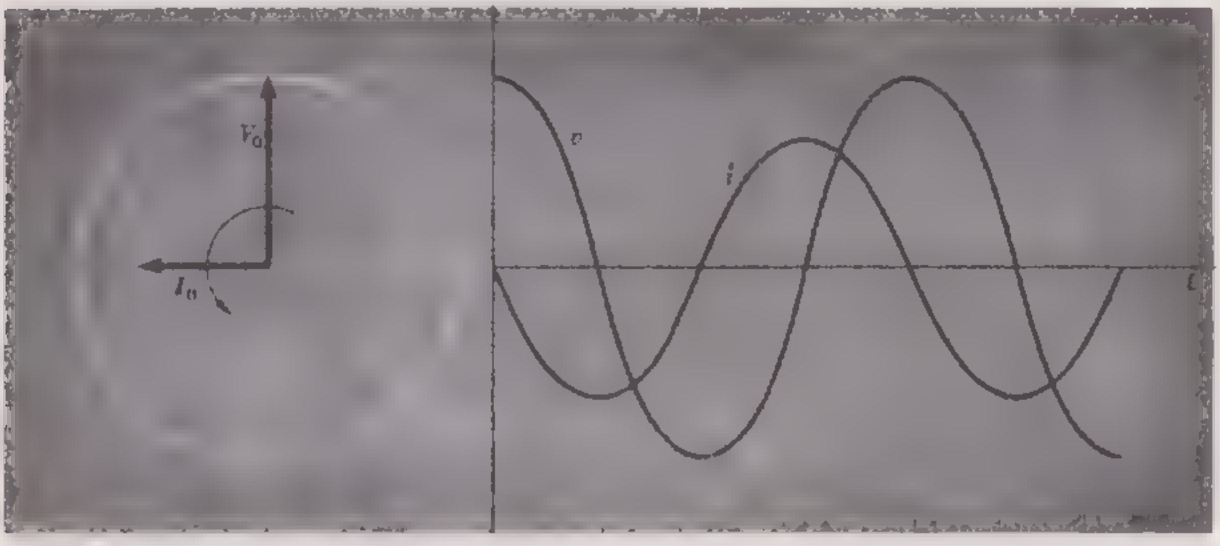


Fig. 10.9 Current-voltage relationship in a capacitance. The current is 90° ahead of the voltage.

90° ahead of the voltage. This is again readily available from the trigonometric relationship,

$$i = -\omega C V_0 \sin \omega t = \omega C V_0 \cos\left(\omega t + \frac{\pi}{2}\right) \tag{10.11}$$

10.4 Series LCR Circuit

We now discuss the complete series circuit of Fig. 10.1. In order to determine the behavior of the circuit, we must first note the important fact that in such a series circuit the instantaneous current in all parts is the same. This follows directly from the continuity equation. Although charge does build up inside the capacitor, outside this there is no storage place, and motion of charge along one conductor implies similar current everywhere else. Even in the capacitor we see that growth of positive charge on one plate requires an equal growth of opposite charge on the other plate, which is equivalent to equal currents on both sides With this principle in

mind, we may obtain the solution to the problem by a simple graphical method.

Figure 10.10 pictures the three generating diagrams shown in Figs. 10.3, 10.5, and 10.7, modified in two ways. The scale has been

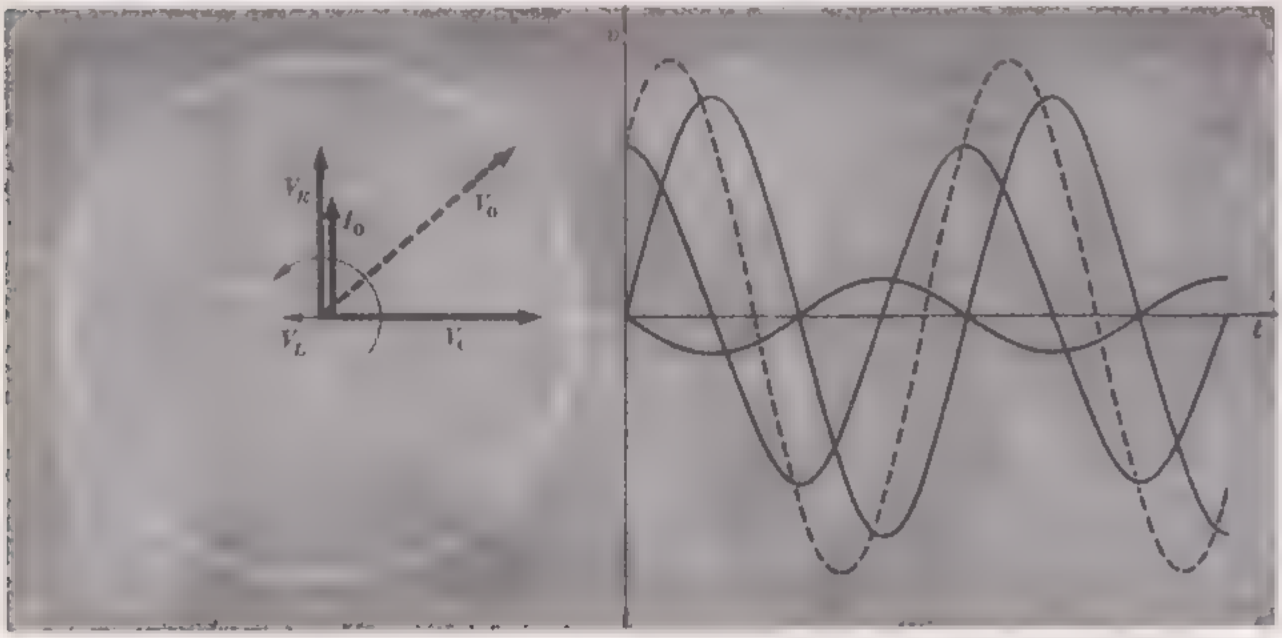


***Fig. 10.10** Generating circle used to obtain graphical solution for a series LCR circuit.*

changed so that the current amplitudes are equal, or $I_R = I_L = I_C$. Also, the diagrams have been rotated relative to each other until the current-generating vectors are parallel. These two modifications are appropriate to the series circuit, in which the current everywhere in the circuits is the same. Having made the amplitude and phase of the current the same in each element, we find that the corresponding voltage-generating vectors V_R, V_L, and V_C now give the relative amplitudes and phases of the sinusoidal voltages across the three elements.

The instantaneous value of the source voltage is equal to the sum of the instantaneous voltages across each element. However, since the sum of sinusoidal voltages of the same frequency is always another sinusoidal voltage, we may express this same fact by saying that the generating vector of the resultant voltage across all three elements is just the vector sum of the individual generating vectors. In order to obtain the amplitude and phase of the generating vector (which is shown dotted in Fig. 10.10), the individual voltage vectors are redrawn as shown in Fig. 10.11 so

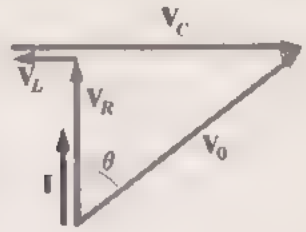


Fig. 10.11 Vector diagram showing vector addition of voltages in a series LCR circuit.

that their vector sum may be obtained. Rotation of the whole diagram at the appropriate rate generates the sinusoidal source voltage of amplitude V_0. Since the current is in phase with V_R, the angle θ is the phase angle between the current and the source voltage V_0. Since our convention is to rotate the diagram counter-clockwise, in our example the current *leads* the source voltage by θ.

When we use the current-voltage relationships for resistance, inductance, and capacitance as worked out above and use the vector diagram as a guide, we can get a quantitative solution for the current in the series LCR circuit as shown in Fig. 10.1. We summarize our earlier results, giving the voltage across each circuit element and its relation to the current through the element:

$$V_R = I_R R \qquad I_R \text{ in phase with } V_R$$

$$V_L = I_L \omega L \qquad I_L \text{ lags } V_L \text{ by } 90^\circ$$

$$V_C = I_C \frac{1}{\omega C} \qquad I_C \text{ leads } V_C \text{ by } 90^\circ$$

As stated earlier, in the series circuit the current everywhere has the same amplitude and phase. Thus we have

$$\mathbf{I}_0 = \mathbf{I}_R = \mathbf{I}_C = \mathbf{I}_L$$

Since in this series circuit the voltage applied to the circuit is the vector sum of the voltage-generating vectors, we can write

$$\mathbf{V}_0 = \mathbf{V}_R + \mathbf{V}_L + \mathbf{V}_C = I_0\left(\mathbf{R} + \omega\mathbf{L} + \frac{1}{\omega\mathbf{C}}\right) \qquad (10.12)$$

Here we have indicated the vector nature of the sum by letting $\mathbf{R}$, $\omega\mathbf{L}$, and $1/\omega\mathbf{C}$ take on the vector nature of the voltage-generating vectors, after dividing out the common term I_0. The vector diagram for the separate terms will now be the same as in Fig. 10.11 except

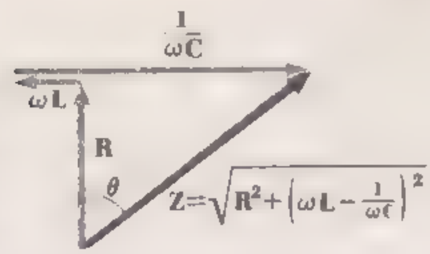


***Fig. 10.12** Vector diagram of impedance vectors. The impedance Z is the vector sum of resistance and reactance terms.*

for scale. This is shown in Fig. 10.12. Because of the right angles involved, the solution of Eq. (10.12) for I_0 becomes

$$I_0 = \frac{V_0}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}} = \frac{V_0}{Z} \tag{10.13}$$

We have introduced the quantity

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \tag{10.14}$$

Z is called the *impedance* of the series circuit. Since

$$Z = \frac{V_0}{I_0} \tag{10.15}$$

the impedance in this a-c circuit takes on the role of resistance in a d-c circuit. Note that whenever the *reactance* terms contribute to Z, the voltage vector V_0 across the circuit is out of phase with the current vector I_0. We have arbitrarily chosen the direction of $L\omega$ in Eq. (10.13) as positive, which makes $1/\omega C$ negative. Since the reactive term $(L\omega - 1/\omega C)$ enters as the square, this choice has no effect on the magnitude of the impedance.

The result given in Eq. (10.13) relates only the amplitude of the a-c current through the circuit to the amplitude of the a-c voltage applied across it. In order to determine the phase relationship between current and applied voltage, we write an expression for the instantaneous current,

$$i = \frac{V_0 \cos(\omega t - \theta)}{\sqrt{R^2 + (L\omega - 1/\omega C)^2}} \tag{10.16}$$

As defined by this equation, θ is the angle by which the current *lags* the voltage applied. We use Fig. 10.11 or 10.12 to obtain the value of θ. We obtain

$$\tan\theta = \frac{\omega L - 1/\omega C}{R} \tag{10.17}$$

As we have chosen the signs of ωL and $1/\omega C$ in the vector diagram, if $(\omega L - 1/\omega C)$ is positive, θ is positive. In our example, θ is negative, corresponding to a current that *leads* the applied voltage.

It is common practice to write the reactive terms in the impedance as follows:

Inductive reactance:

$$\omega L = X_L$$

Capacitative reactance:

$$\frac{1}{\omega C} = X_C$$

Total reactance:

$$X = X_L - X_C$$

Using this nomenclature, the impedance becomes

$$Z = \sqrt{R^2 + X^2} \tag{10.18}$$

We now obtain the same solution to the series LCR circuit of Fig. 10.1 somewhat more formally. If a sinusoidal voltage is applied to the circuit, then the resultant current is also sinusoidal. We may therefore write the following expression for the current and voltage:

$$i = I_0 \sin \omega t \qquad \text{and} \qquad v = V_0 \sin (\omega t - \theta) \tag{10.19}$$

Here we have made only the assumption that the voltage v across the three elements (which must equal the driving voltage of the power source) is sinusoidal, with the same frequency as the sinusoidal current. We are solving to obtain the amplitude and phase angle. The fact that the instantaneous voltages across the three elements add up to the driving voltage can be expressed by

$$RI_0 \sin \omega t + L\omega I_0 \cos \omega t - \frac{I_0}{\omega C} \cos \omega t = V_0 \sin (\omega t - \theta) \tag{10.20}$$

We have added the instantaneous voltages across each circuit element, taking account of phase lag or lead of the voltages by use of the appropriate trigonometric function. Solution of this equation allows us to find the ratio of V_0 to I_0, as well as the value of the phase angle θ between them. Since this equation is true for all

times, we may rewrite it for $\omega t = 0$ and for $\omega t = \pi/2$, to get the equations

$$L\omega I_0 - \frac{I_0}{\omega C} = V_0 \sin(-\theta) = -V_0 \sin\theta \qquad (\omega t = 0)$$

$$RI_0 = V_0 \sin\left(\frac{\pi}{2} - \theta\right) = V_0 \cos\theta \qquad \left(\omega t = \frac{\pi}{2}\right) \tag{10.21}$$

We square each equation and add; using $\sin^2\theta + \cos^2\theta = 1$, we find

$$\left[R^2 + \left(L\omega - \frac{1}{\omega C}\right)^2\right] I_0^2 = V_0^2 \tag{10.22}$$

If we take the square root and solve for I_0, we have again arrived at Eq. (10.13), obtained earlier by graphical solution.

A still more elegant method for solving a-c problems is the method of complex variables. A brief description of the application of this technique is given in Appendix D.

10.5 Parallel LCR Circuit

Another circuit we may examine with ease is the parallel arrangement shown in Fig. 10.13. Here the voltage across each element

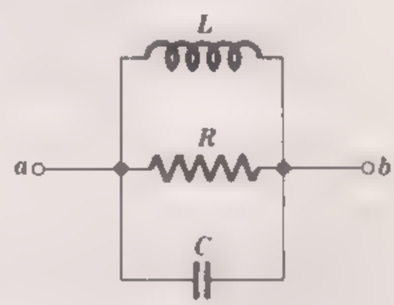


Fig. 10.13 Parallel LCR circuit.

is the same in both phase and amplitude. This then requires that the currents be different. We obtain the individual current amplitudes by use of the following equations:

$$|\mathbf{I}_R| = \frac{V_0}{R}$$

$$|\mathbf{I}_C| = \frac{V_0}{X_C} = \frac{V_0}{1/C\omega} \tag{10.23}$$

$$|\mathbf{I}_L| = \frac{V_0}{X_L} = \frac{V_0}{L\omega}$$

The relative phases of the currents are obtained by noting that in the capacitor the current leads the voltage by 90° and in the inductance it lags by 90°. Thus we may plot a phase diagram for the currents as in Fig. 10.14. We have called I_0 the vector sum of

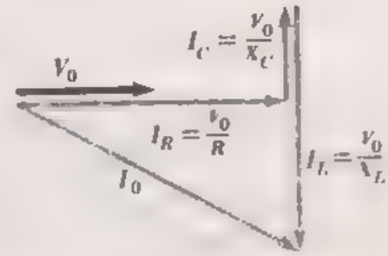


***Fig. 10.14** Vector diagram for parallel LCR circuit, showing vector addition of currents.*

these currents. I_0 is then the current leading to and away from the combination of elements. Since the applied voltage is in phase with the current in the resistance, as we have indicated in the diagram, θ is the phase angle between the resultant current and the applied voltage. We calculate the magnitude of the current using the relationship

$$I_0 = \mathbf{I}_R + \mathbf{I}_C + \mathbf{I}_L = V_0\left(\frac{1}{R} + \frac{1}{X_C} - \frac{1}{X_L}\right) = V_0\,\frac{1}{Z} \tag{10.24}$$

where the vector sum must be taken. Thus, for the parallel circuit,

$$\frac{1}{Z} = \left[\frac{1}{R^2} + \left(\frac{1}{X_C} - \frac{1}{X_L}\right)^2\right]^{1/2} \tag{10.25}$$

Note that the signs of X_C and X_L are reversed in comparison with their use in the series circuit. This results from the equal phases of the voltages in the parallel case compared with equal phases of the currents in the series case. The full expression for the current becomes

$$i = \left[\frac{1}{R^2} + \left(\frac{1}{X_C} - \frac{1}{X_L}\right)^2\right]^{1/2} V_0 \cos(\omega t - \theta) \tag{10.26}$$

where

$$\tan\theta = \frac{1/X_C - 1/X_L}{1/R} = \frac{\omega C - 1/\omega L}{1/R}$$

10.6 Resonance

There is one other aspect of the series circuit of Fig. 10.1, which we now examine. If we look again at the equation for the current in the series circuit,

$$i = \frac{V_0 \cos(\omega t - \theta)}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}} \tag{10.16}$$

we find that for a given set of values of R, L, C, and V_0, the impedance $Z = [R^2 + (\omega L - 1/\omega C)^2]^{1/2}$ varies and hence i varies in both amplitude and phase as the frequency varies. The maximum current occurs for the frequency such that

$$\omega L - \frac{1}{\omega C} = 0 \tag{10.27}$$

If we call this the resonance frequency ω_0, we have

$$\omega_0 = \left(\frac{1}{LC}\right)^{1/2} \tag{10.28}$$

For this particular frequency, i and v are in phase and the current-voltage relationship is as given by Ohm's law. From the diagram of Fig. 10.11 we see that the condition for resonance is that the voltages across L and C are just equal and opposite, so the phase angle is zero. Note also that if the resistance term is very small, the current becomes very large on resonance. Also, the lower the resistance, the more rapidly does the current vary with frequency at frequencies near resonance.

The importance of this resonance type of behavior is very great, since exactly similar effects are found in other fields of physics. One simple example is that of a pendulum in which there is some damping (giving a term equivalent to the resistance term).

The parallel circuit of Fig. 10.13 also shows resonant behavior. In this case the resonance is given by $1/X_C = 1/X_L$, leading again to $\omega_0 = (1/LC)^{1/2}$. Note that on resonance Z is a maximum, giving a minimum current. One way of looking at the situation at resonance is to note that I_0 is just equal to the current through the resistance, while there are large circulating currents through L and C which are of equal magnitudes and in opposite directions and which therefore cancel out to give no contribution to the external current I_0.

10.7 Transients

In any circuit in which energy can be stored, as in a charged capacitor or in an inductance-carrying current, when we apply or remove a voltage suddenly, there is a period during which the circuit adjusts to the new conditions. These temporary effects are called *transients*. We shall discuss a few cases.

The first example is that of an inductance and resistance in series to which we connect a voltage by closing a switch as shown in Fig. 10.15. The resistance here might be simply the minimum

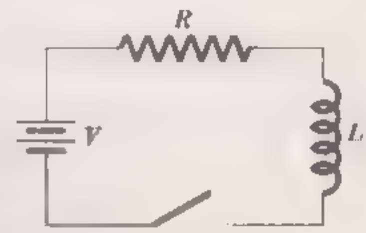


Fig. 10.15 Circuit for showing transient behavior of series RL circuit.

unavoidable resistance in the circuit or it might be a resistor put in on purpose. As soon as the switch is closed, the situation is described by

$$V = Ri + L\frac{di}{dt} \tag{10.29}$$

This is a differential equation, whose particular solution for this situation requires us to put in the initial conditions. In this equation it turns out that the variables can be separated, making its solution very easy. Thus we can write

$$\frac{-R\,di}{V - Ri} = -\frac{R}{L}\,dt$$

where we have multiplied both sides by $-R$ to make the left-hand side a perfect differential. Integration gives

$$\ln (V - Ri) = \frac{-R}{L}t + C \tag{10.30}$$

where we evaluate the constant of integration C from the initial condition that at $t = 0$, when the switch is closed, $i = 0$. This gives $C = \ln V$. We put this in and convert the equation to the exponential form,

$$V - Ri = Ve^{-(R/L)t} \tag{10.31}$$

The current is given by

$$i = \frac{V}{R}(1 - e^{-(R/L)t}) = I(1 - e^{-(R/L)t}) \tag{10.32}$$

We have replaced V/R by I, the *steady-state* value of the current, after a sufficiently long time has passed. A plot of this solution is given in Fig. 10.16. The behavior of the circuit is completely

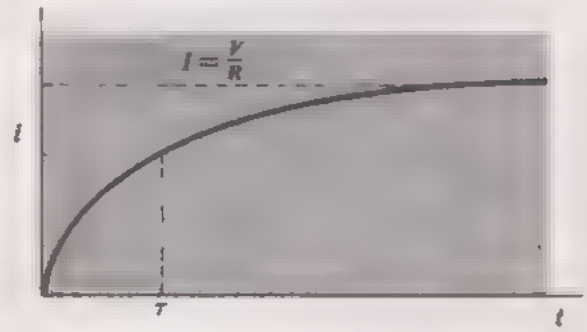


Fig. 10.16 *Time variation of current in an RL circuit.*

determined by the value of R/L, or by its reciprocal $L/R = \tau$, the *characteristic time* of the circuit. With the use of τ as defined, the equation becomes

$$i = I(1 - e^{-t/\tau}) \tag{10.33}$$

τ is the time for the current to build up to $(1 - 1/e)$ or 0.632 of its final value. We see this by letting $t = \tau = L/R$, giving

$$i = I\left(1 - \frac{1}{e}\right) = I(1 - 0.368) = 0.632I$$

We next investigate the effect of suddenly removing the voltage from the circuit, by suddenly changing the switch from A to B as in Fig. 10.17. If we let $t = 0$ at the moment the switch is moved,

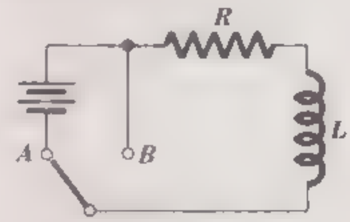


Fig. 10.17 *RL circuit arranged to produce transient by sudden removal of voltage source.*

the equation describing the circuit becomes

$$0 = Ri + L\frac{di}{dt}$$

Separation of variables and integration gives

$$\ln i = -\frac{R}{L}t + C$$

Putting in the initial condition that at $t = 0$, $i = I$, the steady-state current, we get

$$i = Ie^{-(R/L)t} = Ie^{-t/\tau} \tag{10.34}$$

A plot of this equation is shown in Fig. 10.18. Here τ is the time for

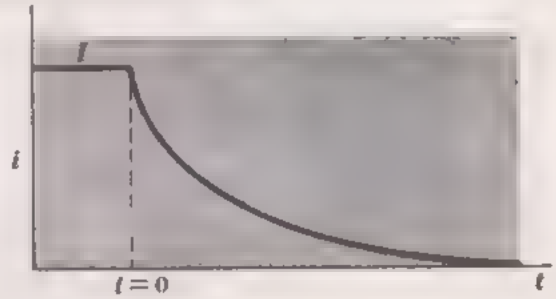


Fig. 10.18 ***Decay of current in RL circuit after voltage source is removed.***

the current to fall to $1/e$ th of its steady-state value.

We next turn to the capacitor circuit shown in Fig. 10.19.

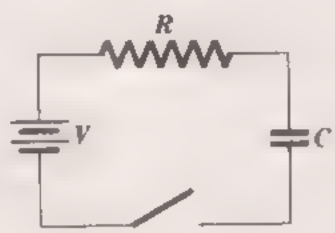


Fig. 10.19 ***Series RC circuit for applying a voltage V suddenly.***

The equation that applies once the switch is closed is

$$V = R\frac{dq}{dt} + \frac{q}{C} \tag{10.35}$$

using dq/dt for the current so as to limit the variables to q and t. Separating variables and integrating, we are led to

$$\ln\left(V - \frac{q}{C}\right) = -\frac{t}{RC} + K$$

When we evaluate the constant of integration K by using the condition that at $t = 0$, $q = 0$, that is, starting with the capacitor uncharged, we obtain

$$V - \frac{q}{C} = Ve^{-t/RC}$$

Solving for q gives

$$q = CV(1 - e^{-t/RC}) = Q_0(1 - e^{-t/\tau}) \tag{10.36}$$

where we write Q_0 for CV, the equilibrium charge on the capacitor, and τ for RC, the time constant here. We can solve for the current by obtaining dq/dt:

$$\frac{dq}{dt} = i = \frac{Q_0}{RC} e^{-t/\tau} = \frac{V}{R} e^{-t/\tau} \tag{10.37}$$

The decay curve is shown in Fig. 10.20. We have also plotted the growth of charge on the capacitor according to Eq. (10.36).

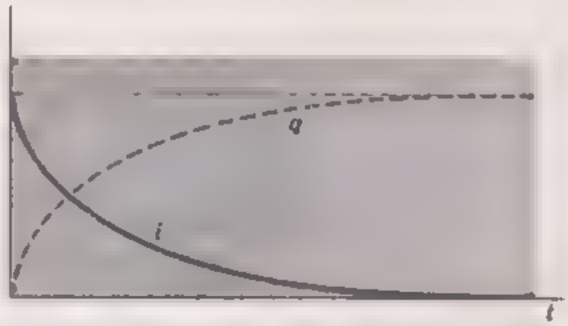


Fig. 10.20 Transient response of a series RC circuit.

Finally, we study the case of the discharge of a capacitor as shown in Fig. 10.21. We start with the capacitor charged to a

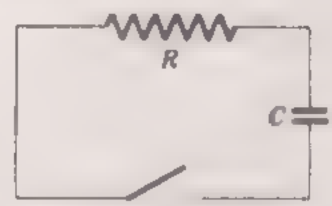


Fig. 10.21 Discharge of a capacitor through a resistance.

potential $V_0 = Q_0/C$. When the switch is closed, the situation is described by

$$\frac{q}{C} + R\frac{dq}{dt} = 0 \tag{10.38}$$

This is the same as Eq. (10.35), except that the battery voltage is zero. Solution by the usual method gives

$$q = Q_0 e^{-t/RC} \tag{10.39}$$

The equation for current is

$$i = \frac{dq}{dt} = \frac{Q_0}{RC} e^{-t/RC} = \frac{V_0}{R} e^{-t/\tau} \tag{10.40}$$

This result is identical with that of Eq. (10.37). However, the current is flowing in a direction to discharge the capacitor in the

latter case, whereas in the former case the current was in the direction to charge the capacitor.

It is possible to make capacitance-resistance circuits having time constants ranging from very short times up to many seconds. As a result, there are many practical uses of such circuits for timing, or time-delay circuits. For example, the most common method of measuring very-high-resistance resistors is to measure the decay rate of charge on a capacitor when the resistor is connected across it. Resistances as high as 10^{12} ohms can be measured in this way. The method, of course, involves the determination of τ, from which R can be determined via $RC = \tau$, if the capacitance is known.

10.8 Filter Circuits

Another field of great practical importance is associated with the response of inductance, capacitance, and resistance circuits to alternating current. This is the design of *filters*. A filter is a circuit that passes certain frequencies of electric signals and attenuates others to a greater or lesser extent. We examine in a qualitative way the behavior of two very simple filter circuits, shown in Fig. 10.22.

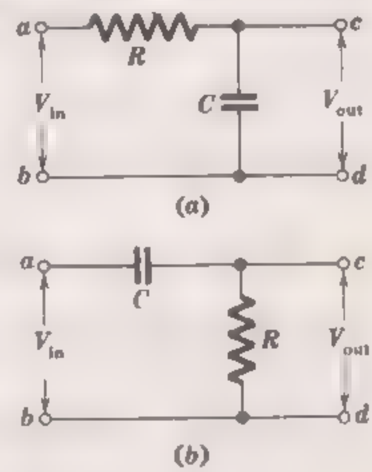


Fig. 10.22 Two simple filter circuits. (a) is a low-pass filter and (b) is a high-pass filter.

The electric signal is brought in at terminals a and b and is taken off at c and d. In (a) we can recognize the RC circuit studied above, where we can think of an a-c voltage attached to terminals a and b. The output is the voltage across the capacitor. The impedance across the resistance is R and across the capacitor is $1/\omega C$. The impedance of the C and R in series is

$$\left[R^2 + \left(\frac{1}{\omega C}\right)^2\right]^{1/2}$$

For a constant input voltage amplitude, we may obtain the output voltage as a function of frequency by taking the ratio

$$V_{out} = V_{in} \frac{1/\omega C}{[R^2 + (1/\omega C)^2]^{1/2}} \tag{10.41}$$

This result is plotted in Fig. 10.23. When $1/\omega C = R$ or $\omega = 1/RC$, the output is $(1/\sqrt{2})V_{in}$. In our simple derivation of Eq.

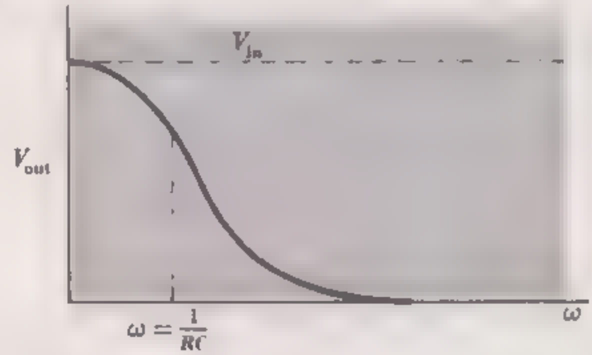


Fig. 10.23 Output of a low-pass filter.

(10.41) we have ignored the possible effects of the impedance across the output. In many applications this will be a resistance or impedance much higher than R or $1/\omega C$ in the frequency range of interest, so its effects can be ignored. This circuit is called a *low-pass filter*, since high-frequency signals (ω a few times $1/RC$) are highly attenuated.

Circuit b of Fig. 10.22 works in the opposite way and is called a ***high-pass filter***. The output voltage is given by

$$V_{out} = V_{in} \frac{R}{[R^2 + (1/\omega C)^2]^{1/2}} \tag{10.42}$$

using the same reasoning as above. The attenuation curve is shown in Fig. 10.24.

In general, a *sharper* cutoff (more rapid variation of attenuation with frequency) can be obtained with more complicated circuits,

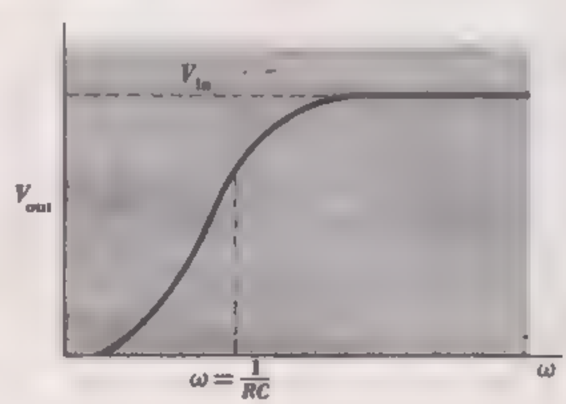


Fig. 10.24 Output of high-pass filter.

usually involving both capacitance and inductance. An inherent feature of considerable importance is that if the attenuation is a function of frequency, it is necessarily true that the phase difference between input and output varies with frequency.

In many applications the signal, or voltage input, to a filter is far from sinusoidal. As discussed in Sec 10.2, however, the behavior of the filter can be understood by dividing the signal into its equivalent sinusoidal components and treating each separately. The method of *Fourier analysis* treats the problem of the relationship between a time-varying function and its sinusoidal components. An example of the selective action of a filter is its effect on music that has been converted into a (complicated) electric signal, as in a radio or phonograph. The ear easily detects the effect of a low-pass filter that removes all sinusoidal components above some fixed frequency, or conversely, the effect of a high-pass filter that removes sinusoidal components below some fixed frequency. More complicated filter networks can be constructed that pass signals only through a certain band of frequencies and stop all others at higher or lower frequencies.

10.9 Power in A-C Circuits

So far we have discussed sinusoidal signals in terms of their frequency and amplitude. In considering power dissipation it is usual to deal with *effective* values of current and voltage. The effective value of an a-c signal is the d-c or constant signal that would develop the same power as the a-c signal. An example that we all use is in calling the a-c voltage in home circuits 110 volts. This is the effective value of the voltage; the a-c amplitude, or peak value, is actually $110 \times \sqrt{2} = 155$ volts. We show how this comes about and also discuss the effect on power dissipation of voltage and current being out of phase, as they usually are in a-c circuits involving capacitance and inductance.

We can develop the relationship between effective and peak values by considering a resistance of value R through which passes a d-c current I_0 for a time t, dissipating an amount of heat $I_0{}^2Rt$. We then replace the direct current by alternating current and wish the heat developed to be given by

$$I_{eff}^2 Rt \qquad \text{joules} \tag{10.43}$$

We want to determine the relationship between I_{eff} defined in this way and the peak value of the alternating current I_0, where $i = I_0 \sin \omega t$.

We know that the instantaneous rate of heating is i^2R watts. Then in a time t, the heat developed will be

$$I_{eff}^2 Rt = \int_0^t i^2R\,dt = \int_0^t I_0{}^2R \sin^2 \omega t\,dt \tag{10.44}$$

When we cancel out R and remove the constant I_0 from inside the integral, we can solve for I_{eff}^2:

$$I_{eff}^2 = I_0{}^2\left(\frac{1}{t}\int_0^t \sin^2 \omega t\,dt\right) = I_0{}^2\,\overline{\sin^2 \omega t} \tag{10.45}$$

where $\overline{\sin^2 \omega t}$ stands for the quantity in parentheses, the average value of $\sin^2 \omega t$. There are numerous ways to show

$$\overline{\sin^2 \omega t} = \tfrac{1}{2} \tag{10.46}$$

A simple graphical way is illustrated in Fig. 10.25, where $\sin^2 \omega t$ is shown compared with $\sin \omega t$. Since the former is also a sinusoidal

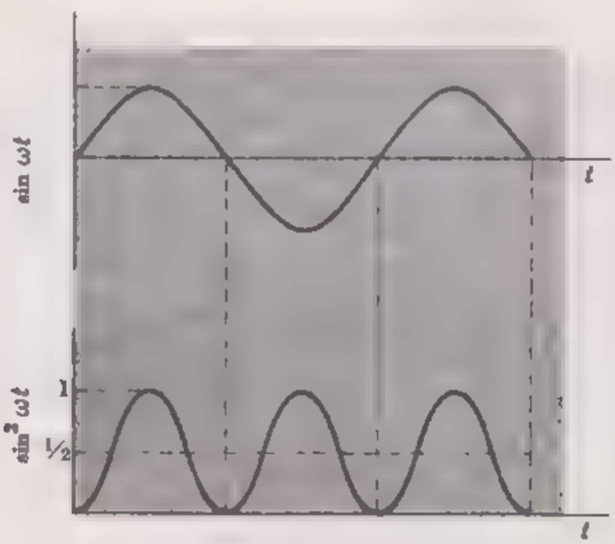


***Fig. 10.25** Graphical demonstration that $\overline{\sin^2 \omega t} = \frac{1}{2}$.*

wave, displaced from the zero axis, we see by symmetry that its average value is ½. This argument can be carried through analytically, by writing the trigonometric relationship

$$\sin^2 \omega t = \tfrac{1}{2}(1 - \cos 2\omega t)$$

Since the average value of $\cos 2\omega t$ is zero, the result of Eq. (10.46) follows at once.

Once we have this result, our problem is solved, since we can then write

$$I_{eff}^2 = \frac{I_0{}^2}{2} \qquad \text{or} \qquad I_{eff} = \frac{1}{\sqrt{2}} I_0 = 0.707\, I_0 \tag{10.47}$$

The same kind of argument gives

$$V_{eff} = \frac{1}{\sqrt{2}} V_0 = 0.707\, V_0 \tag{10.48}$$

The effective values of current and voltage are usually called the *root-mean-square* (rms) values, and these are usually used when a-c power is involved.

We now discuss the effect of a phase angle between I_0 and V_0, as regards power dissipation. We already know that for direct current, the power fed into a circuit that passes a current i and has across it a voltage v is given by $P = iv$. For alternating current, if the current and voltage are in phase and given by $i = I_0 \sin \omega t$ and $v = V_0 \sin \omega t$, the power involved will be $P = iv = I_0 V_0 \sin^2 \omega t$, and the average value will be

$$P = I_0 V_0 \overline{\sin^2 \omega t} = \tfrac{1}{2} I_0 V_0 = I_{eff} V_{eff} \tag{10.49}$$

On the other hand, if a phase angle θ exists between current and voltage, so that

$$\begin{aligned} i &= I_0 \sin \omega t \\ v &= V_0 \sin (\omega t + \theta) \end{aligned} \tag{10.50}$$

then the power is given by

$$P = iv = I_0 V_0 \left[\frac{1}{t} \int_0^t \sin \omega t \sin (\omega t + \theta)\, dt \right] \tag{10.51}$$

We must again evaluate the quantity in brackets, which gives the average value of the quantity inside the integral. Using a familiar trigonometric equation, we rewrite Eq. (10.51) as

$$\begin{aligned} P &= I_0 V_0 \left[\frac{1}{t} \int_0^t \sin \omega t\, (\sin \omega t \cos \theta + \cos \omega t \sin \theta)\, dt \right] \\ &= I_0 V_0 \left[\frac{1}{t} \int_0^t (\sin^2 \omega t \cos \theta + \sin \omega t \cos \omega t \sin \theta)\, dt \right] \end{aligned} \tag{10.52}$$

The second term has an average value of zero since $\sin \omega t \cos \omega t$ is symmetrical about zero and multiplies a constant, $\sin \theta$. Thus we find

$$P = \tfrac{1}{2} I_0 V_0 \cos \theta = I_{eff} V_{eff} \cos \theta \tag{10.53}$$

It follows that the average power dissipation in a pure inductance or a pure capacitance is zero. $\cos \theta$ is called the *power factor*.

10.10 A-C Instruments

Because of the mechanical inertia of the coil of an ordinary galvanometer, it is too much to expect it to respond to a-c currents. In this section we shall examine the principles of a number of a-c measuring instruments that solve this problem by various methods.

Hot-wire ammeter This instrument solves the problem that the torque on a current-carrying coil in a fixed magnetic field reverses when the current reverses by avoiding it completely. It depends instead on the heating of a wire by the current. The heated wire is under slight spring tension so that it elongates when heated, and a pointer is connected so as to measure the extension. As might be expected, this device is not useful for very small a-c currents. Since the heating will be proportional to i^2, if calibrated in terms of i it will give nonlinear response. It can of course be used to measure d-c current, but since galvanometers are usually much more convenient, this is not normally done. If it is calibrated using direct current, it will measure effective or rms current.

Dynamometer This is a galvanometer-type device in which the fixed magnetic field of the usual permanent magnet is replaced by a field produced by current in an auxiliary coil. Both coils are placed in series or parallel so that the current variation is the same in both. Since the magnetic field will now reverse its direction in step with the current in the suspended coil, the torque on the coil keeps the same sign and the deflection is proportional to the average torque. Since the torque is proportional in the first place to the current in the suspended coil and in the second place to the magnetic field, which also depends on the current, the torque depends on i^2. The dynamometer can be used to measure either voltage or current, just as can a galvanometer.

When the two coils are connected separately, one to measure the current to a system and the other to measure the voltage across the system, the dynamometer measures power directly. A schematic diagram is shown in Fig. 10.26. Even if V and I are out of

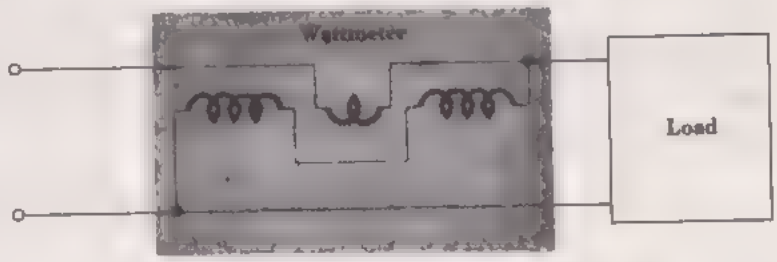


Fig. 10.26 Schematic diagram of a wattmeter.

phase, this meter still gives the correct power since the torque is $iv = I_0V_0 \sin \omega t \sin (\omega t + \theta)$, and, as we have seen, the average value of this is $I_0(V_0/2) \cos \theta$. This type of wattmeter can be used with direct current. Sensitivities of dynamometer devices tend to be low because of the relatively small magnetic fields obtained from available currents compared with the fields of the fixed magnets used in galvanometer movements.

Rectifier galvanometer A rectifying device, usually a semiconductor crystal, is sometimes used to provide direct current, which is then measured by a conventional galvanometer. We discuss rectifiers later. It is sufficient here to remark that the d-c voltage obtained in this way is proportional to the amplitude of the a-c voltage, so the galvanometers give readings proportional to voltage (or current) amplitude.

Cathode-ray oscilloscope (CRO) Although this device comes in a rather different category than the other instruments mentioned, it is in constant use in laboratories for the measurement of a-c and sometimes d-c voltages. In this day of universal familiarity with the CRO in the form of television display tubes, it is probably unnecessary to go into great detail, and in any case this is not the place. A few sentences will suffice. The heart of the CRO is the electron beam, obtained from a hot cathode and focused after electrostatic acceleration on fluorescent material inside the face of the tube. The position of the beam is controlled by either an electrostatic field or a magnetic field produced by a coil. The special advantage of the device is the extremely rapid response of the beam to the voltage applied to the signal electrode. It is common prac-

tice to look at signals in the megacycle frequency range. In addition, the use of an appropriate electronic a-c or d-c amplifier makes possible the easy observation of signals in the range of 10^{-3} volt or less. Although the CRO is not an absolute device, it can be easily calibrated so that its signals can be measured directly in volts. Another great advantage of the CRO is that it has a high impedance input so that it draws very little current and does not upset the circuit it is being used to measure.

10.11 The Transformer

A transformer is a device that allows a change to be made in an a-c supply voltage without appreciable loss of power. Thus a 6-volt lamp may be lit by a 110-volt a-c supply by connecting it through an appropriate transformer, or we can obtain 10,000 volts for an X-ray generator by using an appropriate transformer on the same 110-volt line. If some of the magnetic flux due to current in one coil links another coil so that there is a mutual inductance between the two coils, the pair of coils can be used as a transformer. Usually, however, efforts are made to ensure that as nearly as possible, all the flux of one coil links the other. A very common design is to wind both coils on the same closed ferromagnetic core (Fig. 10.27). As

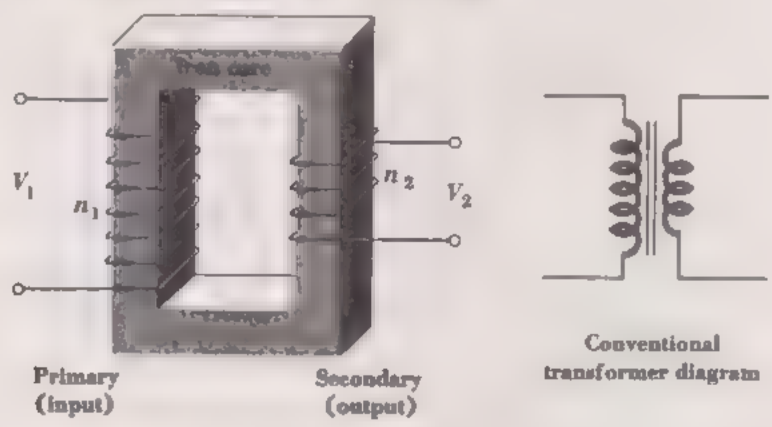


Fig. 10.27 The transformer.

we have mentioned earlier, the flux tends to be confined within the ferromagnetic path so the mutual inductance between the two coils is a maximum.

The ratio of primary to secondary voltage is just the turns ratio of the primary to secondary windings. Suppose the a-c supply voltage to the primary (power input) coil is V_1. If the resistance of the coil is negligible compared with its inductive reactance, an a-c

current i_1 will flow in the coil, causing a varying magnetic flux that is just enough to give an emf equal and opposite to V_1, so we have

$$V_1 = -n_1 \frac{d\Phi}{dt} \tag{10.54}$$

where n_1 is the number of turns in the primary. Now since all the flux links the secondary coil, we have at once for the output voltage

$$V_2 = -n_2 \frac{d\Phi}{dt} \tag{10.55}$$

where $d\Phi/dt$ has the same value as in Eq. (10.54).

Combining these equations, we have

$$\frac{V_1}{V_2} = \frac{n_1}{n_2} \tag{10.56}$$

This is the basic transformer equation that relates the voltage ratio of the primary and secondary windings to the turns ratio of the windings.

We may determine the magnitude and phase of the current in the primary for no load on the secondary by relating the flux Φ to the mmf. As we have seen earlier,

$$\Phi = \frac{\text{mmf}}{\mathcal{R}} \tag{9.54}$$

where $\mathcal{R}$ is the reluctance of the magnetic circuit. When we apply this to the primary winding and assume a sinusoidal primary current, we get

$$\Phi = \frac{n_1 i_1}{\mathcal{R}} = \frac{n_1 I_0 \sin \omega t}{\mathcal{R}} \tag{10.57}$$

Thus the flux varies sinusoidally with time and has the same amplitude and phase as i_1. When we now differentiate to get $d\Phi/dt$ and substitute in Eq. (10.54), we get

$$V_1 = \frac{-\omega n_1{}^2 I_0 \cos \omega t}{\mathcal{R}} \tag{10.58}$$

This shows that i_1 and V_1 are out of phase by 90°, as was to be expected from our general result for a pure inductance. According to

our earlier argument, since the power factor is zero, there is no power dissipation. Also, if the reluctance of the magnetic circuit is small, the current amplitude I_0 is small. Actually, in practical cases there is some power loss because the power factor is not exactly zero. This is the result of the finite resistance of the winding, and of the eddy-current and hysteresis losses in the iron core. In practice, however, these losses are negligible compared with the power consumed by a normal load placed across the secondary winding.

We now examine the change in the situation when a load is connected to the secondary winding, allowing a secondary current i_2 to flow. For simplicity we assume a pure resistive load, so that V_2 and i_2 have the same phase. As a result of the current i_2, an additional magnetic flux is produced in the iron. By Lenz's rule, this flux opposes the original flux due to i_1, so the flux is now given by

$$\Phi = \frac{n_1 i_1}{\mathcal{R}} - \frac{n_2 i_2}{\mathcal{R}} \tag{10.59}$$

But the magnitude of the sinusoidally varying flux must remain fixed if we are to continue to satisfy Eq. (10.54) for the voltage across the primary. This argument uses the fact that Φ and $d\Phi/dt$ are proportional when the frequency is held constant. Thus, as a consequence of introducing i_2, i_1 must increase, and we may write

$$\Phi = \frac{n_1(i_1 + i_1') - n_2 i_2}{\mathcal{R}} \tag{10.60}$$

where i_1' is the increase in i_1 produced by allowing i_2 in the secondary circuit. Comparison of Eq. (10.60) with Eq. (10.57) gives

$$n_1 i_1' = n_2 i_2 \tag{10.61}$$

If we neglect the primary current for no load, i_1' becomes the total primary current and we get the relationship that

$$\frac{i_2}{i_1} = \frac{n_1}{n_2} \tag{10.62}$$

that is, the primary and secondary currents are inversely propor-

tional to the primary and secondary turns. Equations (10.56) and (10.62) lead to

$$V_1 i_1 = V_2 i_2 \tag{10.63}$$

showing that the power input to the transformer is equal to the power output. We are here neglecting the losses, which may amount to a few per cent in ordinary transformers. We shall not make a quantitative study of the losses in transformers but shall discuss briefly the efforts made to reduce eddy-current losses to a minimum. Since these increase with frequency, the problem is most serious at high frequencies. For ordinary 60-cps applications, it is usually sufficient to construct the iron core out of laminations, as we have described in Sec. 8.12. For higher-frequency applications, losses are kept sufficiently low by using finely divided iron particles immersed in an insulating matrix. Finally for very-high-frequency application, ferrites, which are insulating magnetic materials, are used (see Sec. 9.10).

10.12 Generators

We have already discussed in Examples 8.5*a* and 8.8*c* the connection between Faraday's induction law and the operation of an electric generator. Here we add to those discussions a few of the details of construction of practical generators. We saw in Example 8.8*c* how a rotating coil in a magnetic induction field gives a-c power. For d-c operation, two kinds of modification are usual. The first involves the use of a *commutator* to reverse the connections to the coil at the proper time to keep the emf always in one direction, as discussed in connection with the rotating coil used as a motor in Example 8.8*d*. This gives the equivalent of a d-c output with a ripple superimposed. This is carried further, as was discussed in connection with motors, by constructing an armature composed of many separate coils, each of which is connected to the output by the commutator for only that part of the cycle for which its output is nearly maximum. The resulting output has a very much reduced ripple, which can be even further reduced, if necessary, by external filter circuits. The second important modification is to supply the current to the fixed coils, which provide the static magnetic field, from the output of the armature itself. This is the *self-excited* generator.

Figure 10.28 shows two circuits used for applying voltage to the field coils.

The rate at which mechanical energy is converted to electric energy is $\mathcal{E}i_a$, where $\mathcal{E}$ is the emf generated in the armature coils and i_a is the current through them. This neglects the power dissipation

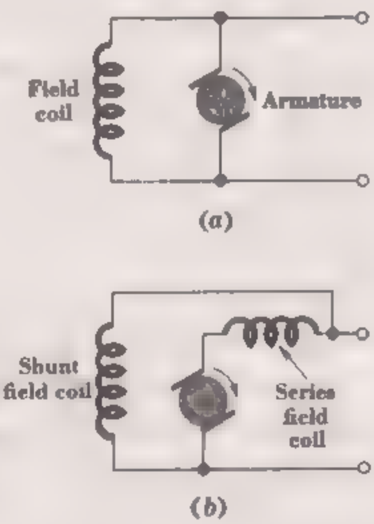


Fig. 10.28 Two types of field-coil excitation circuits for d-c generators: (a) Shunt excitation; (b) compound excitation.

in the resistance of the field coils that maintain the magnetic flux in the generator. In a well-designed generator this loss is a small fraction of the power converted into electric energy.

The usual a-c generator is constructed somewhat differently. The armature coils are placed in fixed positions outside a set of rotating magnetic poles. The poles are made of iron and are kept polarized by a d-c current fed through slip rings to the rotor.

Fig. 10.29 Schematic drawing of a three-phase a-c generator. Rotating magnetic field produces a-c voltages across each pair of terminals, with phase separations of 120°.

In many power applications there are three sets of armature coils connected so that there is an a-c voltage between each of three terminals coming from the coils. The phase of the alternating current between each pair of terminals is different, so that the three lines from the generator deliver what is called *three-phase* power. Figure 10.29 shows a schematic diagram of this arrangement. The

phase angle between each pair of terminals is 120°, as can be inferred from the arrangement. This three-phase power is used for specially designed three-phase motors, which we discuss in the next section.

In all mechanical generators the internal resistance comes from the actual resistance in the coils in which the emf is induced. In the compound-wound d-c generator, an increase in the current to the load results in an increase in the magnetizing current in the series field coil (see Fig. 10.28*b*). The resulting increase in static field can compensate for the otherwise lower voltage output under heavy loads, to give a very low effective internal resistance over a wide range of currents.

We mention briefly two kinds of electrostatic generators. One, the Wimshurst machine, is used now primarily as a low-power, high-voltage machine for lecture demonstrations of the effects of high voltages. It is a mechanized version of the electrophorus apparatus discussed in Sec. 1.5, in which charging is produced by the induction effect. A more sophisticated electrostatic generator is the Van de Graaff machine. In this device an insulating motor-driven traveling belt carries a charge sprayed onto it by a corona discharge from needle points up to a high-voltage terminal where the charge is taken onto a metal sphere surrounding the high-voltage terminal. On sufficiently large models, voltages as high as 12 million volts have been obtained. Most modern machines for use at 1 million volts or more are built inside a pressure jacket to inhibit the loss of charge by corona that tends to occur in air in the high field that develops at the high-voltage terminal. Van de Graaff machines can be made to operate at very constant voltages over considerable periods of time and are therefore of considerable use in nuclear research where thresholds for nuclear processes are to be determined with high accuracy. They are also of great importance for the production of high-energy X rays.

10.13 Motors

In Example 8.5*c* we have seen how magnetic forces on a current-carrying conductor can be used to convert electric to mechanical energy. In Example 8.8*d* we have discussed the back emf in a motor, resulting from the flux change when the moving coil rotates. These ideas cover the fundamentals of motor operation. Here we add a

few more remarks regarding the operation of several common types of motors.

The usual d-c motor consists of a rotating armature containing a number of separate coils connected to the current source by means of a commutator. In miniature motors the magnetic field is provided by permanent magnets, but for most motors the field is produced by electromagnets consisting of iron cores wound with field coils and energized by the external current source.

In the shunt-excited motor, the field coils are connected in parallel with the armature. The current in the armature is given by

$$V_s - \mathcal{E} = i_a R_a$$

where V_s is the source voltage, $\mathcal{E}$ the average back emf due to the motion of the armature coils in the field, and i_a and R_a are the current and resistance in the armature coils. The back emf $\mathcal{E}$ is proportional to Φf, where Φ is the magnetic flux cut by the coils in one revolution and f is the frequency of rotation. The rate at which electric energy is converted into mechanical energy is given by $\mathcal{E}i_a$. This neglects the power dissipated in the resistance of the field coils that maintain the magnetic flux, which is usually small compared with the mechanical work rate. The expression for power conversion is the same as that for conversion from mechanical to electric power in a generator. The rate at which mechanical work is done by the motor is $P = \tau\omega$, where τ is the torque and ω is the angular velocity of the motor. When we equate the electric power going into mechanical work to the mechanical power output, we find

$$\mathcal{E}i_a = \tau\omega \propto \Phi f i_a = \frac{\Phi\omega}{2\pi} i_a$$

or

$$\tau \propto \frac{\Phi}{2\pi} i_a$$

Thus the torque is proportional to the current through the armature windings in a shunt-excited motor.

In a series-excited motor, the field coils and armature windings are in series, so the circuit equation is given by

$$V_s - \mathcal{E} = (R_a + R_f)i_a$$

where R_f is the field-coil resistance. The magnetic flux Φ will be roughly proportional to i_a, and since the torque is proportional to

Φi_a, it will therefore be proportional to i_a^2. This contrast between the torque in shunt- and series-excited d-c motors shows that the series motor has a higher starting torque since current requirements increase only with $\tau^{1/2}$ rather than with τ.

A-c motors can be constructed on the same principles as d-c motors, since the torque depends on Φi_a, and both Φ and i_a change sign at the frequency of the a-c voltage, giving a torque always in the same direction. In a shunt-excited motor, however, there is the difficulty that the phases of the current in the field windings and armature are not the same, since the inductances of the two windings are different. This problem does not exist in series-excited motors, where the currents are necessarily in phase. However, the large inductance of the motor windings limits the current to small values, and this seriously limits the torque available, using ordinary voltage supplies. The iron cores of the field coils must be laminated in order to avoid excessive eddy-current losses.

The most common a-c motor is the induction motor, in which a rotating magnetic field induces currents in a set of copper loops in the rotor. Magnetic forces on these current loops exert a torque on the rotor and cause it to rotate. The rotating field can be produced by feeding three-phase power from a three-phase generator, as described in the preceding section. The motor windings are similar to the windings on the generator. When they are connected to the three out-of-phase lines from the generator, a rotating field is produced in the motor. Another method of producing a rotating field uses a single-phase power supply, but a phase shift in part of the field-coil windings is produced by connecting a capacitance to part of the windings. The intensity of the rotating field is increased by including a laminated iron core in the rotor, thus decreasing the reluctance of the magnetic circuit. This type of motor is not synchronous with the rotating field, since if the rotor moves at the same speed as the field, no currents are induced and there is no torque. A simple induction motor has a low starting torque, since the large induced a-c currents on the stationary rotor induce a large back emf on the field coils, reducing the field-coil current and the magnetic flux. The starting torque can be increased by bringing connections from the rotor loops out through slip rings, to an external resistance that is connected in series while starting.

A synchronous motor can be built by combining regular armature windings that are fed current from the external supply with

some short-circuited loops as in an induction motor. With no load, the rotor comes into synchronism with the rotating field just as a permanent magnet would rotate with the field. When mechanical torque is applied, the rotor falls back in phase and a-c currents are induced in the induction loops. The extra force on the induced-current loops keeps the rotor in synchronism with the rotating field, though behind it in phase. The induction torque increases to a maximum when the phase lag becomes 90°. If the mechanical torque applied produces a lag greater than this, the motor falls out of synchronism.

PROBLEMS

10.1 An inductance of L henrys and a resistance of R ohms are connected in series across a source of a-c voltage given by $V = V_0 \sin \omega t$.

a Find the current i in the circuit and the phase angle between the voltage source and the current. Explain the meaning of phase angle.

b What is the phase angle between the current in the inductance and the current in the resistance?

c What is the phase angle between the voltage across the inductance and the voltage across the resistance?

d Find the amplitude of the voltage V_R across the resistance and the amplitude V_L of the voltage across the inductance.

e Explain why $|V_R| + |V_L| \neq V_0$. Prove that the *vector* sum $\mathbf{V}_R + \mathbf{V}_L = \mathbf{V}_0$.

10.2 In Prob. 10.1, replace the inductance L by a capacitance C farads and answer the same questions.

10.3 *a* Compute the inductive reactance of a 2-henry inductor at 60 cps.

b Compute the capacitative reactance of a 50-μf capacitor at 60 cps.

c At what frequency would the reactances of these two elements be equal in magnitude?

d Plot a qualitative curve of the reactance of each as a function of frequency.

10.4 A tightly wound circular coil of area A has N turns of wire and rotates about its diameter, which is perpendicular to a uniform magnetic induction field B. It has a frequency of rotation $\omega = 2\pi f$. The coil has a self-inductance L and is connected to an external resistance R.

a Write the expression for the magnetic induction flux linkage through the coil as a function of the angle $\theta = \omega t$ between the plane of the coil and the direction of B.

b Write the expression for the emf induced in the coil.

c What is the amplitude of this emf?

d Find the current in the coil.

e Find the phase angle between the induced emf and the current and explain what this means.

f What is the phase angle between the voltage across the external resistance and the current through it?

g What is the phase angle between the voltage across the coil and the current through it?

10.5 A 60-cps voltage of amplitude 120 volts is placed across a pure resistance of 20 ohms. What is the rms value of the applied voltage? Find the peak current, average current, rms current, and power dissipation.

10.6 A pure inductance of 0.2 henry is placed across a 400-cps voltage source of amplitude 120 volts. Find the peak current, average current, rms current, and power dissipation.

10.7 A 60-cps voltage is placed across a 400-ohm resistance in series with a capacitor of unknown capacitance. An a-c voltmeter across the source reads 80 volts and an a-c ammeter (of negligible resistance) in the circuit reads 0.1 amp. Calculate:

a The impedance of the circuit

b The power dissipation in the circuit

c The value of the capacitance

10.8 Show that the average value of a sinusoidal alternating current during the positive half-cycle is $2I_0/\pi$, where I_0 is the current amplitude.

10.9 Calculate the rms value of a voltage given by

$$V = 50 \sin \omega t + 20 \sin 2\omega t$$

10.10 A series circuit as shown in Fig. P10.10 has resistance R_1, R_2 ohms, capacitance C farads, and inductance L henrys. An a-c voltage of amplitude V_0 volts is placed across the circuit.

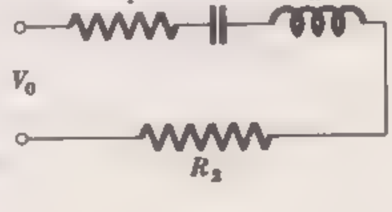


Fig. P10.10

a For what frequency will the current in the circuit be a maximum?

b What will be the value of the maximum current?

c For what frequencies will the current be one-half its maximum value?

d For what frequencies will the current be a minimum?

10.11 A parallel LCR circuit as shown in Fig. P10.11 has an a-c voltage of amplitude V_0 applied.

a For what frequency is the current in the capacitance C the same magnitude as the current in the inductance L?

b For what frequency is the current through the a-c source a maximum? A minimum?

c For what frequencies is the current through the a-c source twice the minimum value?

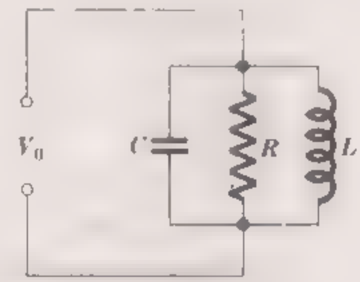


Fig. P10.11

10.12 The switch in the circuit shown in Fig. P10.12 is closed at $t = 0$.

a Make a qualitative plot of the potential V_R across the resistance R as a function of time. On the same time scale plot V_L, the potential across the inductance L. Add the two curves to get the sum $V_R + V_L$.

b At what time does $V_R = V_L$?

c What is the characteristic time of this circuit?

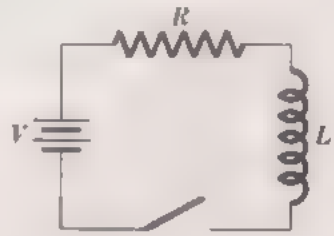


Fig. P10.12

10.13 Suppose that after the switch in the circuit shown in Fig. P10.12 has been closed for a time long compared with the characteristic time of the circuit, the switch is almost instantaneously opened. Make a qualitative plot of the voltages V_R and V_L versus time. Can you see any limitation on the speed with which the current can be reduced to zero?

10.14 *a* In the circuit shown in Fig. P10.14, find the time after the switch is closed when the voltage $V_R = V_C$.

b What is the maximum current that can flow in the circuit, and when does this maximum occur?

c What is the maximum charge on the capacitance and when does this occur?

d Show that the current in the circuit drops down to half its original value at the same time as the charge on the capacitor reaches half its final value. What is the time after closing the the switch that these two events occur?

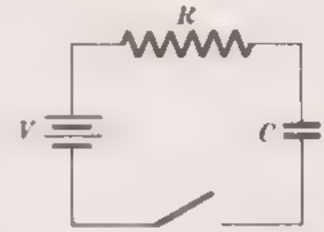


Fig. P10.14

10.15 An induction coil of inductance 10 henrys and resistance 100 ohms is connected to a 20-volt battery. What resistance must be connected in parallel with the coil in order to prevent the voltage across the coil from rising above 100 volts when the battery circuit is suddenly opened? What is the initial rate of decrease of current in the inductance?

10.16 Design a simple high-pass filter such that the voltage output is one-half the voltage input at a frequency of 400 cps (see Fig. 10.22). Assume the filter circuit feeds into a much higher impedance circuit.

10.17 A transformer, as shown in Fig. P10.17, has n_1 primary turns and n_2 secondary turns. Sixty per cent of the magnetic induction flux due to current in the primary goes through the shorting branch A. Find the ratio of primary to secondary voltage with no load on the secondary.

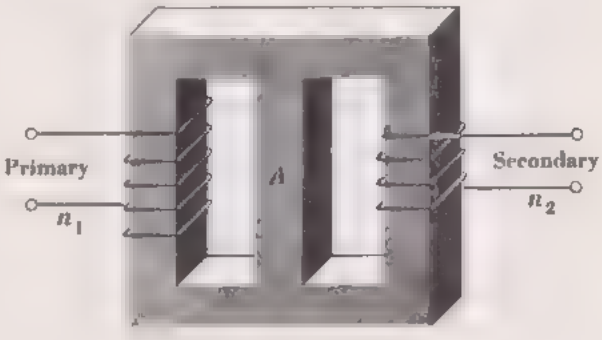
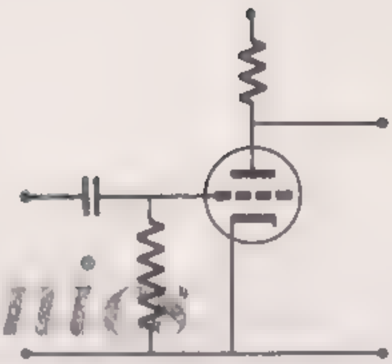


Fig. P10.17

ELEVEN

Electronics



11.1 Introduction

In this chapter we touch briefly on the basic characteristics of three devices of primary importance in practical electronics. These are the vacuum tube, the transistor, and a special high-frequency tube, the klystron. The vacuum tube and transistor are basically similar in function in that the current through them can be controlled by an electric signal. The entire field of electronics depends on the control characteristics of these devices. The many special types of tubes and transistors allow for a multitude of electronic circuits that perform very diverse functions. It is appropriate here to discuss only their basic design and a few typical applications. The klystron is discussed as one example of devices that operate at very much higher frequencies than conventional vacuum tubes. They are the most important sources of microwave-frequency signals and are becoming of major importance in the communication field. The availability of electric power at microwave frequencies has allowed the development of many new fields of investigation in experimental physics.

11.2 The Vacuum Tube

The principal function of the vacuum tube is to control the flow of current in a circuit. This control is possible with negligible expenditure of power. As a result, the vacuum tube can be used to amplify an a-c signal or to operate as an oscillator in appropriate circuits and thereby produce an a-c signal from a d-c source of power. In a vacuum tube, thermionic emission from a hot *cathode* in a vacuum provides a constant source of electrons that carry a negative current to a *plate* or *anode*, held at a positive potential with respect to the cathode. Control of this current is obtained through the operation of one or more *grids* whose potential is controlled by external circuits. We begin our considerations with a discussion of the relationship between the vacuum current and the potential applied between cathode and plate.

The electron current in a vacuum is limited by the rate at which electrons can be emitted by the cathode. This depends on the temperature of the cathode, but in practice the controlling factor is the *space charge* in the cathode region. The effect can be explained qualitatively as follows: In the absence of electron emission, there is a field at the cathode owing to the potential difference applied between plate and cathode. When electrons are emitted by the cathode, this field accelerates the electrons toward the cathode and there is a distribution of electrons in the cathode region. Their presence alters the field distribution and lowers the field near the cathode, which in turn allows more electrons to collect in the cathode region. This process continues until the field at the cathode surface actually reverses. In steady state, the kinetic energy of the emitted electrons just overcomes the negative field, and there exists a steady-state *space-charge-limited* current. A simple quantitative analysis of this situation, which we shall not develop here, shows that the current is proportional to $V^{3/2}$, where V is the plate-to-cathode potential difference. To a good approximation this law holds for any electrode geometry. Notice that the vacuum tube does not behave as an ohmic resistance, since the current is *not* proportional to the applied voltage.

We first investigate the properties of a vacuum-tube *diode*, which has only a cathode and plate and which is used primarily as a *rectifier*. Figure 11.1 shows a simple circuit in which a diode is used as a *half-wave rectifier*. In this circuit an a-c voltage input is converted to direct current. We have shown the cathode heated by a

heater filament connected to a low-voltage external supply, usually a transformer. This supplies the heat to cause the cathode to emit electrons. The cathode is usually coated with a metal oxide that lowers the temperature required to give electron emission. Whenever the plate is positive with respect to the cathode, electrons from

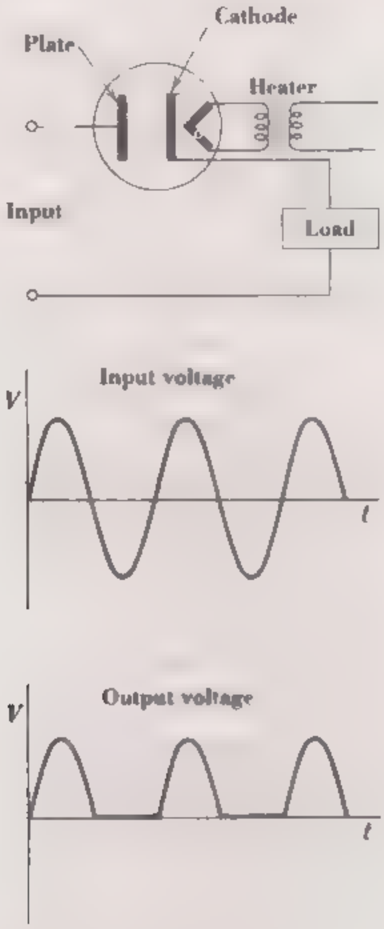


Fig. 11.1 Half-wave rectifier.

the cathode are attracted to the plate, causing a positive current from plate to cathode. Thus on the half-cycle of the a-c voltage during which the plate is positive, current flows through the tube. On the negative half-cycle, however, no current can flow, since electrons will be repelled by the negative plate voltage. The result of this action is the output signal shown, in which the negative half-cycles are missing. The average value of the half-wave-rectified voltage output is no longer zero, and if a suitable filter is added, a net d-c voltage is obtained. The simplest filter would be obtained by connecting a capacitor across the output.

A full-wave rectifier circuit using a center-tapped transformer

is shown in Fig. 11.2. The two tubes required are often placed in the same vacuum envelope and use a common cathode. Also shown is a more complicated filter circuit such as is used for minimizing the remaining a-c ripple. On alternate half-cycles the current passes through that tube which has a positive potential,

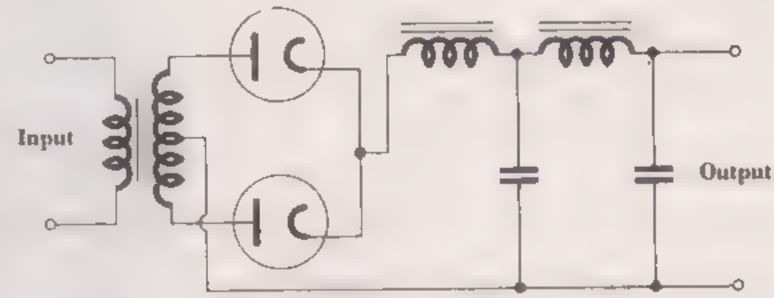


***Fig. 11.2** Full-wave rectifier, using center-tapped transformer, with filter circuit in output. Cathode heaters are not shown.*

while the other tube ceases to conduct. The heater connections are omitted in the diagram. Such a rectifier circuit is the usual source of d-c voltage needed for amplifiers, oscillators, and other electronic devices.

In a *triode*, shown schematically in Fig. 11.3, a single control grid is placed between cathode and plate. The control grid is a

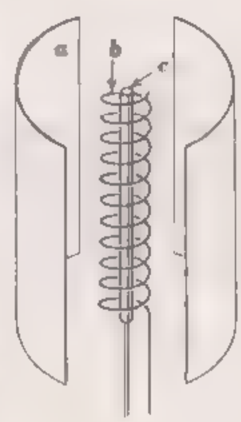


***Fig. 11.3** Triode vacuum tube, showing (a) heated cathode, (b) spirally wound grid, and (c) plate.*

very open network of wire through which most of the electrons can pass on their way to the positive plate. However, if a negative potential (relative to the cathode) is applied to the grid, the field set up will modify the space-charge-limited current and tend to prevent electrons from reaching the plate, thus reducing the current. The magnitude of the plate current can thus be controlled from zero to a maximum value by varying the grid potential. Since practically no current goes to the grid, this can be done with a very small expenditure of electric power. Figure 11.4 shows a plot of plate current versus grid voltage when the plate voltage is held constant. This plot shows that some current goes to the plate even when the grid is quite negative with respect to the cathode.

This is due partly to the open structure of the grid, which allows some lines of force to penetrate through the grid from plate to

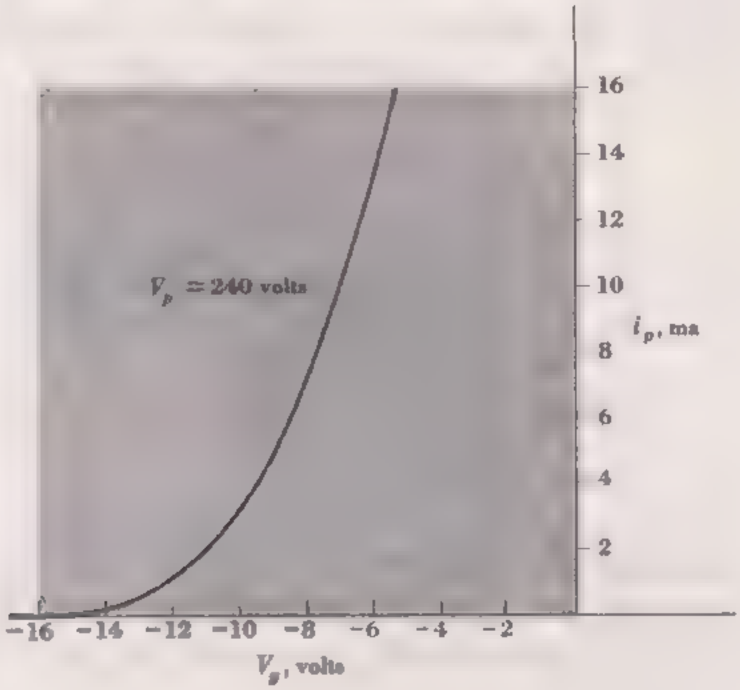


***Fig. 11.4** Plate current versus grid voltage in a triode. Plate voltage held constant.*

cathode even when the grid is negative. Another factor involved is the finite kinetic energy of the electrons evaporated from the cathode, which enables them to overcome a small retarding potential. Figure 11.5 shows the more conventional plot of the

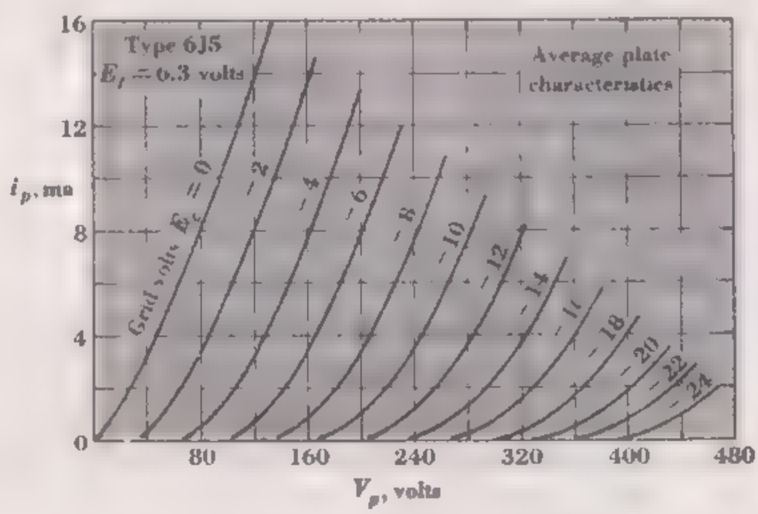


***Fig. 11.5** Triode plate current versus plate voltage for various grid voltages. (From RCA Receiving Tube Manual, Copyright 1950, Technical Series RC16, Fig. 92CM-4771T, page 147.)*

characteristics of a triode, from which Fig. 11.4 was obtained. The current through the tube depends on both the grid and plate potential.

Figure 11.6 shows a typical circuit in which a triode is used as an a-c amplifier. We use this to give a qualitative picture of the process of amplification. Let us imagine an a-c signal (not necessarily sinusoidal), having an average amplitude of, say, about 0.1 volt. If the reactance $1/C_g\omega$ is very small compared with the grid

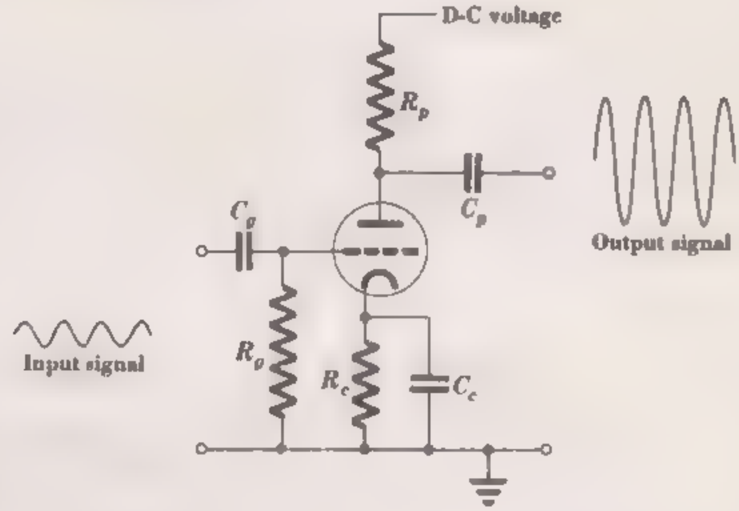


***Fig. 11.6** A triode used in a typical amplifier circuit.*

resistance R_g, most of the voltage will appear across R_g and on the grid of the tube. The cathode resistance R_c and the plate resistance R_p will have been chosen to put the tube in its operating range (where the current is sensitive to the grid voltage). The capacitor C_c across the cathode resistor is large enough to hold the cathode at a nearly constant voltage despite the rapid variation of the current through the tube due to the a-c signal. That is, the time constant of the R_cC_c circuit is long compared with the period of one cycle at the signal frequency. The small grid signal causes a variation in the tube current; this causes the voltage drop across the plate resistance R_p to vary, and the resulting a-c signal at the output becomes perhaps one hundred times the input signal, say, 10 volts. By placing a series of tube stages one after the other, large amplifications can be obtained.

A self-excited oscillator is shown in Fig. 11.7. A *tank* circuit, consisting of an inductance and a capacitance and having a natural frequency of oscillation, feeds back a small amount of signal to the input grid. If the phase of this signal is correct, the consequent variation of current in the tube reinforces the oscillations in the tank circuit and they become self-maintaining. There are many alternative oscillator designs, but in all cases the function of the

tube is to feed back enough energy to the oscillating circuit to make up for the losses that would otherwise cause it to stop oscillating.

More complicated tubes have extra electrodes that modify the characteristics of the tubes for special applications. Discussion of such tubes as tetrodes and pentodes is omitted here, though

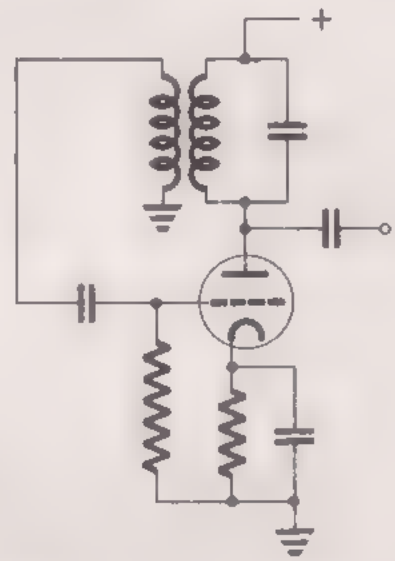


Fig. 11.7 ***A triode used in a self-excited oscillator circuit.***

they are of great importance in electronic devices. We also omit discussions of such special-purpose tubes as beam-power tubes and of gas-filled tubes such as thyratrons and voltage regulators. The interested student may readily inform himself of the characteristics of the multitude of types of tubes by studying one of the several tube manuals readily available. Our purpose here has been to give only the briefest qualitative description of the principles involved.

11.3 Transistors

The transistor was first developed in 1948, and it has already become of great importance as a substitute for vacuum tubes in many applications. Since transistors involve important electrical properties of semiconductors, we discuss them briefly here. Their principal advantage over tubes is their small size and their lack of need for a hot cathode. We begin by giving a brief and necessarily crude description of a semiconductor, the material of which transistors, as well as crystal diodes, are made.

In order to describe the properties of a semiconductor, we are again faced with the difficulty that some of its most important

properties can only be explained on a quantum-mechanical basis. As in our treatment of metals, we shall simply describe the results of the theory without giving arguments or proofs. We begin by showing the contrast between the behavior of metals and semiconductors It was shown earlier that in a metal the conduction electrons completely filled the allowed energy levels up to a certain maximum value W_F. There are allowed states above this limit, but except for a few thermally excited electrons, these higher energy states are unoccupied. These empty levels, however, are essential for conduction to occur. Suppose, for instance, that in a material there is a band of allowed energies completely filled by electrons and that above this band of filled states there is an energy region for which there are no allowed states. In such a material no current flows when an external electric field is applied. In order that electrons be accelerated in an applied field, and hence contribute to conduction, they must be able to move to higher energy levels. But since *all* levels are occupied and since not more than one electron is allowed in each state (counting two electrons of opposite spin as being in different states), the only process allowed is an exchange of position between two electrons in different states. This process does not change the total momentum of the electron system and therefore does not lead to a conductivity. Such a material is classed as an insulator. A semiconductor differs from this by virtue of having another band of allowed energy levels whose states are normally unoccupied; this is separated from the lower filled band by an energy gap in which electronic states are not normally allowed. This contrast is shown schematically in Fig. 11.8.

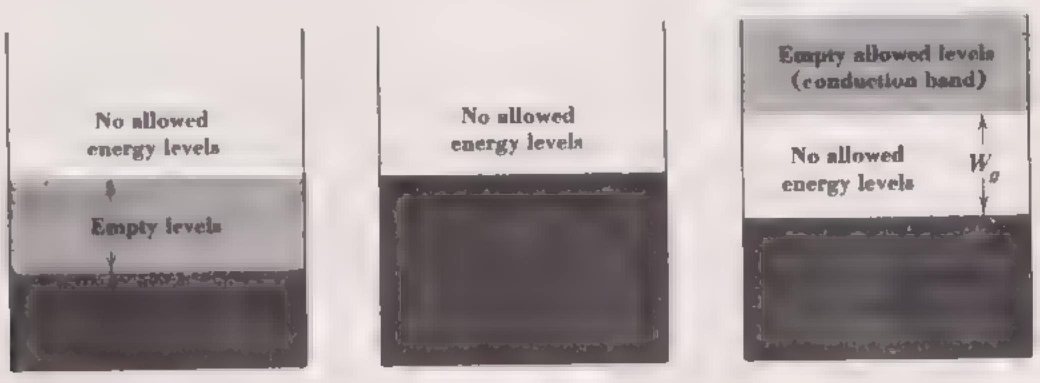


Fig. 11.8 Energy-band diagrams for (a) a metal, (b) an insulator, and (c) a semiconductor. Energy in plotted vertically.

The behavior of semiconductors is easily explained on this model. If the forbidden gap W_g is large enough, ordinary thermal energies or external fields are not able to alter the population in the filled band and the material cannot conduct. However, if there is some mechanism that can give an energy as great as W_g to electrons at the top of the filled band, these can be moved up into some of the upper levels, where they are free to conduct by moving through the various allowed upper levels under the force of an external field. In addition, each electron excited into the upper *conduction* band leaves behind an empty allowed state. Since electrons in the filled band can now alter their energy by exchanging position with the empty level left behind, the presence of such empty levels in the lower or *valence* band allows conductions to occur there also. We may concentrate our attention on the effective motion of these empty levels and thus think of this kind of conduction as being caused by *holes* in the energy distribution.

We describe the course of events either in terms of the transfer of electrons from one allowed state to another or in terms of actual motion of electrons from one position in space to another. This is permissible since when the energy state of an electron changes, its momentum and velocity change; this produces a change in the total momentum of the band of electrons that leads to a net shifting of electron positions in space.

When a semiconductor is raised to a high enough temperature, some electrons are thermally excited into the conduction band, leaving holes behind, and conduction occurs through the motion of electrons in the conduction band and holes in the valence band. Another way to produce conduction is to irradiate the material with radiation whose quanta have energies in excess of W_g, in which case this energy can excite electrons into the conduction band, again leaving holes behind. Holes behave as positively charged carriers since their effective motion in space is opposite to the net motion of the electrons with which they exchange position The excited state so produced is not an equilibrium state, and there is a tendency for electrons to fall back into the valence band and combine with holes.

Our description thus far has been of an ideal semiconductor having no impurity atoms, which can alter the behavior of semiconductors in very important ways. We describe these effects using Ge or Si as an example. These elements are in Group IV in the

periodic table, and both form crystal structures like that of carbon in diamond. The special feature of the diamond lattice for our purposes is that each atom has four nearest neighbors. Each atom shares its four valence electrons with its four nearest neighbors, giving a double electron bond between all nearest neighbor atoms, as shown schematically in Fig. 11.9. This model accounts for all the

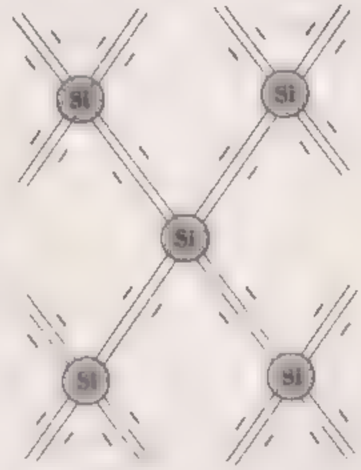


***Fig. 11.9** Schematic representation of atoms in a silicon crystal, showing the four neighboring atoms next to each atom. The true arrangement is three-dimensional. Bonding electron pairs are shown.*

valence electrons as being localized between pairs of atoms. The energy states of these electrons comprise the filled valence band of the material. It is the excitation of one of these electrons into a state in which it is no longer bound and the vacant state left behind that constitute the conduction-band electron and the hole in the valence band, respectively.

Consider the effect of two kinds of impurity atoms which, when they are introduced into the crystals, replace Si or Ge atoms in the lattice. First consider Group V atoms such as P, As, or Sb. Since there are five valence electrons rather than four as in Ge or Si, one more electron will be required in the vicinity of the atom to maintain neutrality than is needed to fill the bonding positions. This extra electron is bound to the impurity atom only by the coulomb force of attraction to the nucleus, and is held much less tightly than the electrons that participate in the electron bonding between atoms. As a result, the *extra* electrons contributed by Group V impurity atoms can be raised into the conduction band by the absorption of much smaller energies than those required for raising electrons from the valence band into the conduction band. The situation is shown schematically in Fig. 11.10. The electronic levels produced by the impurity atoms are shown near the top of the forbidden energy gap. At low temperature, electrons occupy

these levels, though relatively low energies of excitation lift them into the conduction band. These extra energy levels contributed by the impurity atoms are called *donor* levels, and material into which this type of atom has been introduced is called *n-type* (for negative) material.

When, instead, impurity atoms from Group III of the periodic table, such as Al, Ga, or In, are introduced into sites normally filled by the Si or Ge atoms, the situation is reversed. That is,

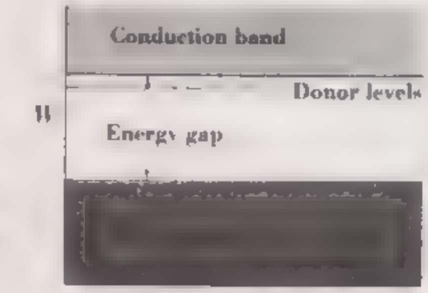


***Fig. 11.10** Energy-band diagram of an n-type semiconductor.*

electrical neutrality requires one less than the four electrons needed to fill all the bonding positions between the impurity atom and its neighbors. It takes relatively little energy to remove a nearby bonding electron and place it in the bonding position that is lacking an electron at the impurity atom. In effect, such a removal of a bonding electron amounts to the production of a hole. This situation is shown in Fig. 11.11. When the *acceptor* levels of

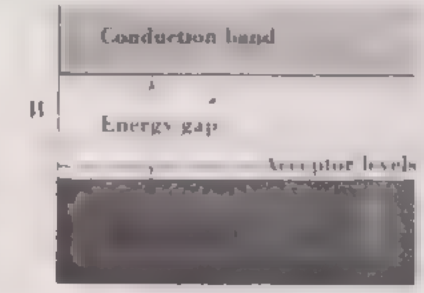


***Fig. 11.11** Energy-band diagram of a p-type semiconductor.*

the impurity atoms are filled with excited electrons, holes are left in the valence band and conductivity can occur via the holes. This kind of material is called *p-type* (positive).

One kind of transistor utilizes sandwich construction in which a thin layer of p-type material separates two n-type regions. This is called an n-p-n junction transistor, and we now discuss its operation. Figure 11.12 shows the design of the transistor. The contacts on the surfaces of the *emitter*, *base*, and *collector* regions allow electrons to flow between the material and the external circuit. The p region

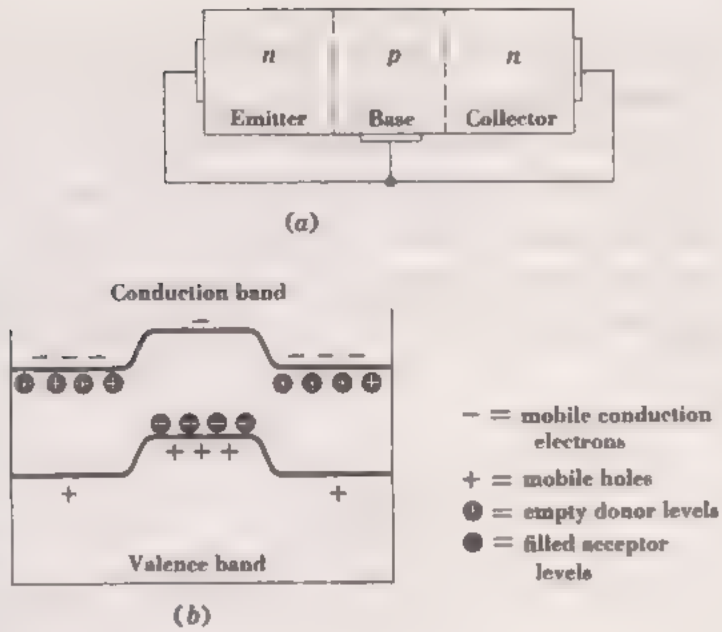
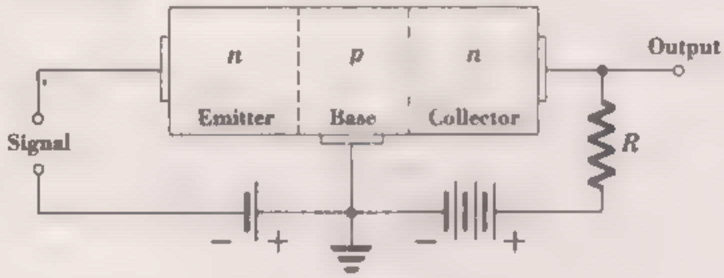


***Fig. 11.12** (a) n-p-n junction transistor; (b) electron energies in an n-p-n junction transistor with no applied voltage.*

is actually very thin. In order to understand the operation of this device, we must realize two facts. The first is that we are dealing with a dynamic equilibrium in both n- and p-type materials. Thus in n-type materials, electrons are constantly being thermally excited into the conduction band from donor levels while other electrons are returning to the donor levels. In p-type materials, there is a flux of electrons up into the acceptor levels and from these levels down into holes in the valence band. Occupied donor levels are electrically neutral, while when unoccupied there is an associated positive charge. Occupied acceptor levels are negatively charged, while when unoccupied they are neutral. The second fact is that when an n-type and a p-type material come into contact, there is a net flow of electrons from the n to the p material and a similar flow of holes from p to n material until there are fields set up that bring this net flow of charge to equilibrium. This flow results from the diffusion motion of holes in the valence band and of electrons in the conduction band. Diffusion brings about a net charge transfer because nearly all the conduction electrons are generated in the n-type region and nearly all the holes are generated in the p-type region.

Figure 11.12 shows the kind of potential distribution that results when all three regions are connected together externally. There are conduction electrons and positively charged empty donor levels in the n-type regions, and negatively charged filled acceptor levels and mobile holes in the p-type material. The net charge

transfer due to diffusion produces a dipole layer at each junction. Charge diffusion has occurred, as indicated in the diagram by the presence of one conduction electron in the p-type region and one hole in the n-type region. There is constant recombination of holes and electrons in each region, and this is just counterbalanced by the diffusion currents. In the diagram given, higher electron energies are upward, while higher hole energies are downward.



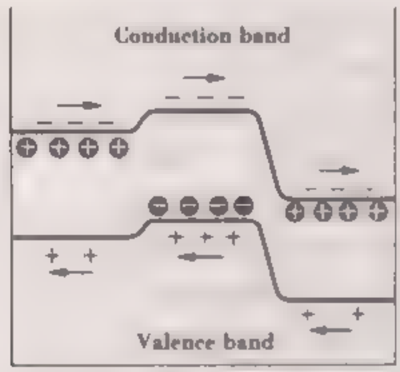


***Fig. 11.13** Electron energies in an n-p-n transistor with voltages applied.*

Figure 11.13 shows the modifications in this situation when external potentials are applied. Under the influence of the applied fields more electrons are able to diffuse to the right from the n region to the base; once they enter the base, they are forced into the collector region by the applied field. Most of the applied voltage appears across the transition layers between n and p regions because these have much higher resistivity than do the pure n and p regions. A signal placed on the emitter electrode controls the flow of electrons in a manner similar to the action of a control grid in a triode. At the same time there is a counterflow of holes to the left; these combine with electrons at the emitter electrode to allow an electron current to flow. Since both flow direction and sign of charge are opposite for holes compared with electrons, both types of carriers give the same direction of current flow. The amplified signal is obtained through the variation of current through the resistance R. This is only one of several types of

transistor design, but it suffices to show the general principles involved. In all cases, advantage is taken of the thermal supply of charge carriers in a *doped* or impurity-containing semiconductor.

Another very important semiconductor device is the crystal diode. This has a much longer history than the transistor, although the most recent designs now use *n-p* junctions as in transistors. In all cases crystal diodes pass current more easily in one direction than in the other as a result of the dipole layer resulting from thermal diffusion in a semiconductor. We omit further discussion of these very important devices.

11.4 The Klystron

During World War II* there was a very great development of equipment utilizing very-high-frequency a-c circuits having frequencies from 1,000 to, say, 24,000 megacycles/sec (Mc/sec). The development of power sources and detectors of power in this microwave frequency range made radar possible and has also opened up many new fields in experimental physics. In this section we limit discussion to a brief description of one of the most important methods of generating these very high frequencies, the klystron tube.

There are a number of reasons why a conventional vacuum tube becomes unsatisfactory as an oscillator or amplifier at frequencies above a few hundred megacycles per second. A major difficulty lies in the small but unavoidable capacitances between control electrodes in the tube. Since the impedance of these capacitances depends on $1/\omega C$, when frequencies get too high the capacitance has the effect of shorting out the high-frequency voltage. A second difficulty arises when the transit time for electrons to travel from cathode to plate becomes comparable to a period of the high-frequency signal. Under these conditions the rapidly varying electron density in the electron beam induces high-frequency currents in the grid that prevent effective control of the current. This difficulty limits conventional tubes to frequencies below about 100 Mc/sec. Present-day schemes for getting around this difficulty actually take advantage of transit-time effects. As an example we shall discuss the *reflex klystron*, which is shown schematically in Fig. 11.14. A hollow metal cavity in the shape of a toroid, which at the high frequencies involved acts like a resonant *LCR* circuit, is attached to a pair of grids through which there travels a beam of

electrons from the cathode. A negatively charged repeller electrode causes the electrons to be turned around and return through the grid system in the opposite direction. We first discuss the operation of the cavity and grid system. If we consider the cavity with grids as an *LCR* resonant circuit, the two parallel grid areas act as the principal capacitance and the toroidal walls of the cavity act as the inductance. At the very high frequencies involved, lumped circuit

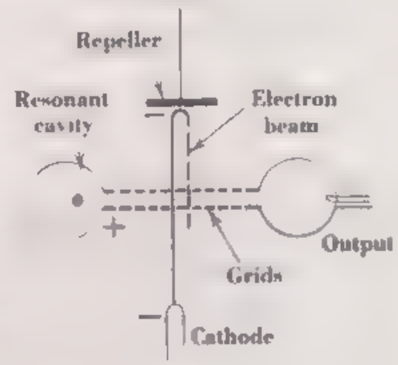


Fig. 11.14 *A reflex klystron (schematic).*

elements have little meaning, and it is usually more convenient to think of an oscillating electromagnetic wave inside the metal boundary. (Electromagnetic waves are discussed in Chap. 12.) Thus the relative potential between the two parallel grids oscillates as the result of currents flowing around the cavity from one grid to the other, as caused by the oscillating electromagnetic field. The problem is to feed back energy into this oscillating field to compensate for losses and for the energy taken out at the output. This is accomplished by means of the electron beam, which gets its energy from the d-c voltage drop between cathode and grid. Let us imagine that the cavity is oscillating as the electron beam passes through the pair of grids. Depending on the phase of the grid-voltage oscillation, some electrons are accelerated when they pass through, while others are slowed down. This introduces velocity differences into the electron beam, which in turn result in a *bunching* of electrons at certain positions along the beam. By appropriate adjustment of the repeller voltage, this bunching can be made to occur just as the beam returns through the grids, giving a series of pulses of electrons with the time interval between pulses equal to the period of oscillation of the grid voltage that caused the bunching to occur. If, further, the electron bunches pass through the gap between grids at a phase such that they are retarded by the field,

the energy loss they suffer goes into the electromagnetic field. This is the required mechanism by which d-c power is converted into microwave oscillations. Other microwave oscillators include magnetrons and traveling-wave tubes, which we shall not discuss.

PROBLEMS

11.1 In Fig. P11.1, a tube with characteristics as shown in Fig. 11.5 is connected to a 240-volt d-c supply through a plate resistance R_p = 30,000 ohms. We wish to investigate the effect of a change in the grid voltage V_g on the plate voltage V_p.

a Using Fig. 11.5, plot the *load line* for this triode for the given values of applied voltage and R_p. To do this, we plot two or more points on Fig. 11.5 that give the voltage V_p across the tube for various plate currents. Thus for zero current, there will be no voltage drop across the plate resistor, so $V_p = 240$ volts. Similarly, when $i_p = 8$ ma, the voltage across the tube will be $V_p = 240 - iR_p = 240 - (8 \times 10^{-3} \times 30{,}000) = 0$. The straight line between these two points is called the load line. For the circuit shown, this tube will operate along this line.

b Using the load line, find graphically the change in plate voltage when the grid voltage is changed from -6 to -4 volts.

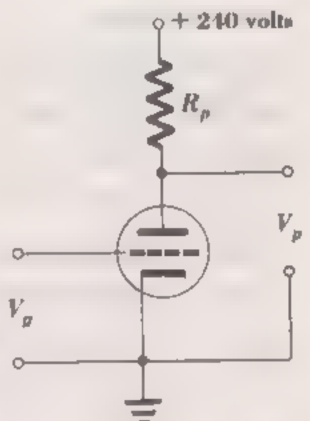


Fig. P11.1

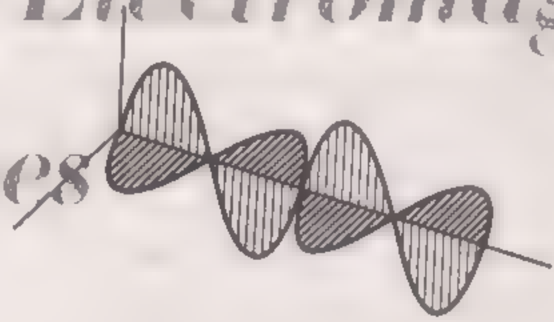
11.2 Give a qualitative discussion to explain why, in an amplifier circuit such as shown in Fig. 11.6, there is a phase shift of 180° between input and output signals.

11.3 Contrast briefly the temperature-dependence of the resistivity of a metal and that of a semiconductor.

11.4 Explain why an n-type and a p-type semiconductor, brought into contact, suffer a change in electrostatic potential. Which becomes more positive and why?

TWELVE

Maxwell's Equations and Electromagnetic Waves



12.1 Introduction

The fundamental experimental facts about electricity and magnetism studied so far include all but one of the basic ideas of the modern theory of this subject. In this chapter we indicate the revolutionary work of Maxwell, published in 1865, which took the individual and seemingly unconnected phenomena of electricity and magnetism and brought them together into a coherent and unified theory. Maxwell's theoretical work showed up a missing principle and led to his postulation of energy transmission by *electromagnetic waves*. Maxwell's discovery that electromagnetic waves should travel with a velocity extremely close to that known experimentally for the velocity of light led him to conclude that light itself is an electromagnetic wave, as had been suggested earlier by Faraday.

12.2 The General Theory of Electromagnetism

We begin by rewriting the experimentally determined equations discussed earlier in a more generalized form equivalent to the

formulation by Maxwell. For simplicity we consider the vacuum case, though it is easy enough to modify the discussion to take account of the presence of dielectric and magnetic materials.

We begin by writing the equation that comes directly from Faraday's law of induction,

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\mu_0 \int \frac{d\mathbf{H}}{dt} \cdot d\mathbf{S} \tag{12.1}$$

The left-hand term is the line integral of the electric field around any closed path in space. This amounts to the work per unit charge to take a charge around the path, and it is therefore the induced emf $\mathcal{E}$. According to Faraday's law, this emf is the negative of the rate of change with time of the magnetic flux through the area surrounded by the closed path. But this is just the meaning of the right-hand term, since $\int \mathbf{B} \cdot d\mathbf{S} = \int \mu_0 \mathbf{H} \cdot d\mathbf{S} = \Phi$, and therefore if $\mathbf{H}$ is changing with time, we can write $d\Phi/dt = -\mu_0 \int d\mathbf{H}/dt \cdot d\mathbf{S}$. The integral is to be taken over the area surrounded by the closed path we choose. Qualitatively this simply says that an electric field is set up in a region of space where the magnetic flux is changing with time. Thus Faraday's law gives a connection between a varying $\mathbf{H}$ and a resulting $\mathbf{E}$. Nothing new has been added to the previous statement of Faraday's law, but the present formulation emphasizes the fact that this is a property of space and does not depend on the presence of a conducting loop of wire.

The next two equations are

$$\int_{cs} \mathbf{D} \cdot d\mathbf{S} = 0 \tag{12.2}$$

and

$$\int_{cs} \mathbf{B} \cdot d\mathbf{S} = 0 \tag{12.3}$$

These are the equations which state that lines of **D** and **B** are continuous. They tell us that the net flux of lines of **D** or **B** is zero through any arbitrary volume. The net flux of **D** is zero because we assume a vacuum. As discussed earlier, these equations require an inverse-square law of force, since only in this case can fields be described in terms of continuous lines of force.

We have now written three of the four fundamental relationships of Maxwell. In the next section we see the way in which Maxwell was led to the fourth and final relation, involving the new concept of *displacement current.*

12.3 Displacement Current

In the process of unifying electromagnetic theory, Maxwell discovered the final relationship that leads to the possibility of electromagnetic waves. This relationship is the converse of Faraday's law, as given in Eq. (12.1). Maxwell showed that just as a varying magnetic field gives rise to an electric field, a varying electric field gives rise to a magnetic field. The argument shown here is similar to the one given by Maxwell.

We begin by examining the generalized form of Ampère's law for the production of a magnetic field by a current:

$$\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0 i \qquad \text{or} \qquad \oint \mathbf{H}\cdot d\mathbf{l} = i \tag{6.14}$$

An equivalent form is

$$\oint \mathbf{H}\cdot d\mathbf{l} = \int \mathbf{j}\cdot d\mathbf{S} \tag{12.4}$$

Here the line integral is taken around any closed path, and the surface integral is taken over *any* surface bounded by the path of the line integral. The surface integral of the current density $\mathbf{j}$ is the total current threading the path of the line integral. When we choose the path so that $\mathbf{H}$ is constant and parallel to the path, $\mathbf{H}$ can be evaluated, as we have seen earlier. Figure 12.1 illustrates the idea that *any* surface bounded by the path of the line integral gives the same result. In the case of a wire carrying the current, we know this is true since the surface integral is always equal to i

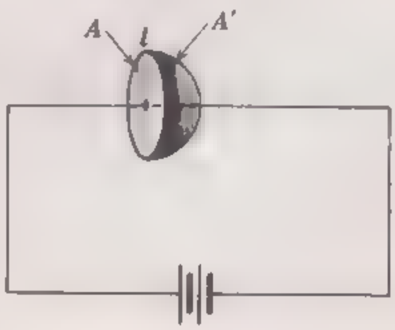


Fig. 12.1 Two areas A and A′ bounded by the same closed line l.

as long as the wire threads the path. The figure shows one plane surface A and one hemispherical surface A', both bounded by the same line l.

We now examine the situation when the battery is replaced by an a-c source and a capacitor is included in the circuit, as shown in Fig. 12.2. An a-c current passes through a capacitor, as we have

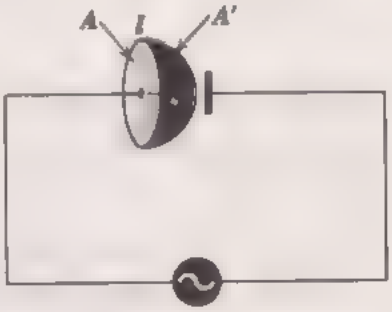


***Fig. 12.2** The integral $\oint \mathbf{H}\cdot d\mathbf{l}$ over the area A is i; it is zero over A' unless displacement current is taken into account.*

seen earlier, though no actual charge is transferred between the plates. Maxwell's reasoning involved taking the two areas A and A', bounded by the same path l. He pointed out that according to Eq. (12.4), when the surface integral is taken over A, the integral $\oint \mathbf{H}\cdot d\mathbf{l} = i$, but if it is taken over A', the integral equals 0. This contradictory situation can be remedied only if we postulate, with Maxwell, an additional term in Eq. (12.4) in which the changing electric field occurring in the capacitor takes the place of the real current as a producer of the magnetic field. We thus arrive at Maxwell's displacement current and remove the double meaning of the mathematics of Eq. (12.4) in the a-c case.

If we use a simple parallel-plate capacitor, it is easy to arrive at the required relationship. Since the current is the rate at which charge accumulates on the capacitor plates (since charge is conserved), we can write

$$i = \int \mathbf{j}\cdot d\mathbf{S} \tag{12.5}$$

But $i = dq/dt = A\, d\sigma/dt$, or $j = d\sigma/dt$. We have already seen that between the plates of a parallel-plate capacitor, $D = \sigma$, so $d\sigma/dt$ can be replaced by dD/dt. Then Eq. (12.4) becomes

$$\oint \mathbf{H}\cdot d\mathbf{l} = \int \frac{d\mathbf{D}}{dt}\cdot d\mathbf{S} \tag{12.6}$$

in the region between the capacitor plates. That is, the displacement current dD/dt acts like a current density j in that it produces a region of magnetic field. The complete expression involving both terms, the magnetic effect of the real current in the wire and the magnetic effect of the change in D, can be written as

$$\oint \mathbf{H}\cdot d\mathbf{l} = \int \mathbf{j}\cdot d\mathbf{S} + \int \frac{d\mathbf{D}}{dt}\cdot d\mathbf{S} \tag{12.7}$$

For most situations involving conducting bodies, the new term involving the displacement current is trivial compared with the current term at low frequencies. In a vacuum it is the only term.

Since we show below that this new concept of magnetic field generation by displacement current is necessary to account for electromagnetic waves, we may consider the proved existence of such waves as the final proof of the validity of Maxwell's reasoning.

The development of the idea that light is an electromagnetic radiation is one of the most dramatic in the history of physics. We shall quote below from the introduction to the paper [1] in which Maxwell announced the electromagnetic theory of light. In the introduction, Maxwell summarizes the results of his mathematical investigations, describing eight relationships, which he calls the general equations of the electromagnetic field. These are essentially equivalent to the four relationships now called Maxwell's equations. Maxwell then goes on to discuss some of the implications of his results in the following paragraphs:

> The general equations are next applied to the case of a magnetic disturbance propagated through a nonconducting field, and it is shown that the only disturbances which can be so propagated are those which are transverse to the direction of propagation, and that the velocity of propagation is the velocity v, found from experiments such as those of Weber, which expresses the number of electrostatic units of electricity which are contained in one electromagnetic unit.
>
> This velocity is so nearly that of light, that it seems we have strong reason to conclude that light itself (including radiant heat, and other radiations if any) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws. If so, the agreement between the elasticity of the medium as calculated from the rapid alternations of luminous

[1] James Clerk Maxwell, A Dynamical Theory of the Electromagnetic Field, *Phil. Trans. Roy. Soc. London*, **155**:459 (1865).

vibrations, and as found by the slow processes of electrical experiments, shows how perfect and regular the elastic properties of the medium must be when not encumbered with any matter denser than air. If the same character of the elasticity is retained in dense transparent bodies, it appears that the square of the index of refraction is equal to the product of the specific dielectric capacity and the specific magnetic capacity. Conducting media are shown to absorb such radiations rapidly, and therefore to be generally opaque.

The conception of the propagation of transverse magnetic disturbances to the exclusion of normal ones is distinctly set forth by Professor Faraday in his "Thoughts on Ray Vibrations." The electromagnetic theory of light, as proposed by him, is the same in substance as that which I have begun to develop in this paper, except that in 1846 there were no data to calculate the velocity of propagation.

12.4 *The Physics and Mathematics of Waves*

Our next task is to see how Maxwell's general formulation of electromagnetism leads to self-propagating waves. Before doing this, however, it is useful to review the simple theory of traveling waves. We discuss the simple situation of wave propagation on a stretched string.

Figure 12.3 shows a section of a long string tied at both ends, under a tension T. A section of the string has been pulled trans-

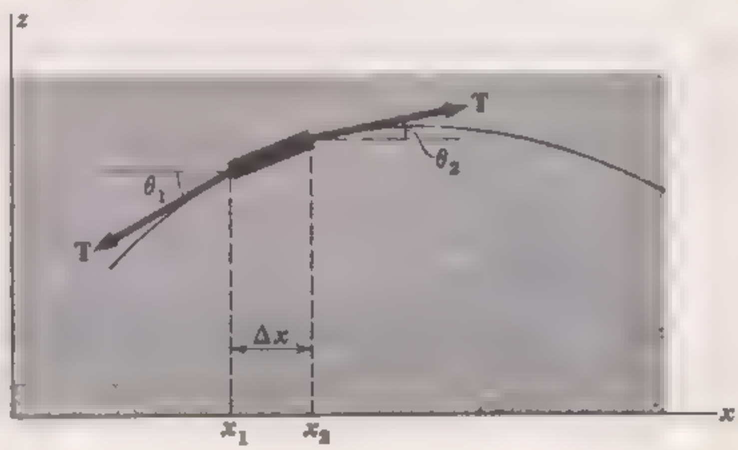


***Fig. 12.3** Element on a stretched string, used for developing traveling-wave equation.*

versely in the direction z and then released. As a result, a wave of displacement travels down the string in the x direction. We concentrate our attention on a small segment Δx of the string, in order to work out the way in which the wave moves along the string. The motion of the segment is obtained by applying Newton's

second law, so we first work out the external forces acting on it as a result of the tension in the string.

The diagram shows that the components of force along the x direction cancel out for small displacement. Let ΔF_x be the net force in the x direction. Then we find

$$\Delta F_x = T \cos \theta_2 - T \cos \theta_1 \sim 0$$

since for small displacements $\cos \theta_2 \sim \cos \theta_1$. As a result of this, the motion of the segment is only along the z direction. The net force along the z direction is given by

$$\Delta F_z = T \sin \theta_2 - T \sin \theta_1 \tag{12.8}$$

For small displacements,

$$\sin \theta_1 \sim \tan \theta_1 = s_1$$

$$\sin \theta_2 \sim \tan \theta_2 = s_2$$

where s_1 and s_2 are the slopes of the line at x_1 and x_2. That is,

$$s = \frac{dz}{dx} \tag{12.9}$$

For small Δx, we may write

$$s_2 = s_1 + \frac{ds}{dx} \Delta x \qquad \text{or} \qquad s_2 - s_1 = \frac{ds}{dx} \Delta x \tag{12.10}$$

Here it is assumed that the rate of change of slope is linear with x over the small distance Δx. With this approximation, we may rewrite Eq. (12.8) as

$$\Delta F_z = T \frac{ds}{dx} \Delta x \tag{12.11}$$

We now apply Newton's law to the segment, noting that its mass can be written as $\Delta m = \rho(x_2 - x_1) = \rho \, \Delta x$, where ρ is the mass density of the string:

$$\Delta F_z = \Delta m \frac{dv_z}{dt} \qquad \text{or} \qquad T \frac{ds}{dx} \Delta x = \rho \, \Delta x \frac{dv_z}{dt}$$

When Δx is canceled in the last equation, we find

$$\rho \frac{dv_z}{dt} = T \frac{ds}{dx} \tag{12.12}$$

an equation that relates acceleration to the rate of change of slope along x. We take the derivative of Eq. (12.12) with respect to time to get

$$\rho \frac{d^2 v_z}{dt^2} = T \frac{d^2 s}{dx\, dt} \tag{12.13}$$

Finally, we simplify this by using the following relationships:

$$v_z = \frac{dz}{dt} \longrightarrow \frac{dv_z}{dx} = \frac{d^2 z}{dx\, dt}$$

$$s = \frac{dz}{dx} \longrightarrow \frac{ds}{dt} = \frac{d^2 z}{dx\, dt}$$

$$\therefore \frac{ds}{dt} = \frac{dv_z}{dx} \tag{12.14}$$

When we differentiate Eq. (12.14) with respect to x, we find

$$\frac{d^2 s}{dt\, dx} = \frac{d^2 v_z}{dx^2} \tag{12.15}$$

Substitution in Eq. (12.13) then gives

$$\rho \frac{d^2 v_z}{dt^2} = T \frac{d^2 v_z}{dx^2} \tag{12.16}$$

This final equation is the differential equation for a transverse traveling wave on a string. It is essentially the equivalent of $F = ma$, giving the transverse acceleration in terms of the effective force, $T\, d^2 v_z/dx^2$.

We can verify at once that wave motion along the string is implied by seeing that the differential equation is satisfied by an equation for a transverse traveling wave. We take

$$v_z = v_0 \cos \frac{2\pi}{\lambda}(x - ut) \tag{12.17}$$

which is a transverse-velocity sinusoidal wave of wavelength λ traveling with a velocity u along the positive x direction. Substitution of the appropriate derivatives of this equation in (12.16) shows that the velocity of travel is given by

$$u = \sqrt{\frac{T}{\rho}}$$

It is now demonstrated explicitly that Eq. (12.17) represents a wave traveling along the x direction, in this case in the positive direction, with a velocity u and a wavelength λ. The wavelength is determined by holding t constant and noting that if we move from some position x_0 to a new position $x_0 + \lambda$, the argument of the cosine goes from $2\pi(x_0 - ut)/\lambda$ to $2\pi(x_0 + \lambda - ut)/\lambda = 2\pi(x_0 - ut)/\lambda + 2\pi$. Since the cosine is repetitive with a period 2π, it has the same value at these two positions, so V_z at these positions has the same value. This result holds for all positions $x_0 + n\lambda$, where n is an integer, so λ is the wavelength of the disturbance. The velocity of the wave is determined by choosing a point x along the wave at a time t and requiring that this point move along x with a velocity such that $(x - ut) =$ constant. Under this condition the cosine remains constant and the point moves with the velocity of the disturbance. If the time is increased by dt, the position x must increase by dx so that

$$x + dx - u(t + dt) = (x - ut)$$

This requires that $dx - u\,dt = 0$, or $dx/dt = u$. We have thus shown that u is the wave velocity. For a wave traveling in the negative x direction, Eq. (12.17) is written as

$$v_z = v_0 \cos \frac{2\pi}{\lambda} (x + vt)$$

We may also use Eq. (12.17) to deduce the relationship between frequency f, wavelength λ, and u, the velocity of wave propagation. For this purpose we stay at one position along the x axis (at $x = 0$ for convenience). The argument of the cosine then becomes $2\pi ut/\lambda$. The disturbance that propagates along x then results in a periodic variation in v_z at the point $x = 0$ we have chosen. If we now let time advance from t to $t + \lambda/u$, the argument of the cosine becomes $2\pi ut/\lambda + 2\pi$, which gives the same value to v_z as for t. Thus λ/u is the *period* or u/λ the *frequency* f of the disturbance, and we have

$$\lambda f = u \tag{12.18}$$

We could equally well have made this discussion in terms of an equation for the *displacement* of the string z, from its equilibrium position,

$$z = z_0 \sin \frac{2\pi}{\lambda} (x - ut)$$

Use of sine or cosine is of course arbitrary and depends only on the choice of initial conditions. With very little more trouble it is possible to show that *any* function of $(x \pm ut)$ is a solution of the differential equation for a traveling wave, so we are by no means limited to sinusoidal waves. With these results for a simple mechanical system we now look at the problem of electromagnetic waves.

12.5 Electromagnetic Waves

For convenience, we rewrite here Maxwell's equations for a vacuum before showing how they lead to traveling waves.

$$\oint \mathbf{E}\cdot d\mathbf{l} = -\mu_0 \int \frac{d\mathbf{H}}{dt} \cdot d\mathbf{S} \tag{12.1}$$

$$\oint \mathbf{H}\cdot d\mathbf{l} = \int \frac{d\mathbf{D}}{dt} \cdot d\mathbf{S} \tag{12.6}$$

$$\int_{CS} \mathbf{D}\cdot d\mathbf{S} = 0 \tag{12.2}$$

$$\int_{CS} \mathbf{B}\cdot d\mathbf{S} = 0 \tag{12.3}$$

In order to simplify this discussion, we consider only the case of plane waves, with the direction of propagation along the x axis. In Sec. 12.6 we show that for a plane wave, the electric and magnetic fields involved in the wave are directed transverse to the direction of travel (in the yz plane) In the meantime we assume this and further specialize our model by assuming a *polarized* wave, one in which the electric field is along only the y direction This field E_y depends only on x and t. We shall see below that $\mathbf{H}$ is similarly limited and is in the z direction, perpendicular to $\mathbf{E}$. Figure 12.4 is a sketch of the $\mathbf{E}$ vectors (*not* lines of force) that might occur.

Fig. 12.4 **E vectors in xy plane, showing variation of E with x.**

Only the variation with x is shown. The variation with time could be shown by making a number of sketches in which the position of the wave along the x axis was different for each time chosen. A sinusoidal variation is drawn here, but this is not required. The vectors are shown only along the x axis of our coordinate system. In this plane wave the same **E** fields exist for *all* positions in the yz plane having a given value of x. Our demonstration involves the use of the first two Maxwell equations to show that such a postulated time and space variation of **E** gives rise to a similar time and space variation of **H** (but at right angles to **E**) and that this **H** variation acts back to cause the postulated variation in **E**. Thus, once such a wave is initiated, it is self-propagating.

Figure 12.5 is used to show the application of Eq. (12.1) to the plane **E** wave, postulated to be moving along the x direction. A

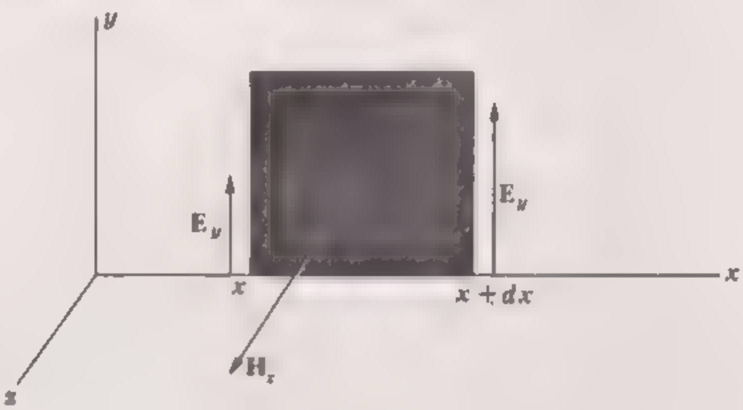


***Fig. 12.5** Generation of H_z by a varying E_y.*

convenient closed path is drawn in the xy plane, around which we shall take the line integral of **E**. This is equated through (12.1) to the rate of change of flux of **H** through the plane bounded by the path of the line integral. Only the vertical parts of the line integral contribute since **E** is in the y direction, so that $\mathbf{E}\cdot d\mathbf{x} = 0$. If we go around in a counterclockwise direction, the line integral around the path chosen becomes

$$\oint \mathbf{E}\cdot d\mathbf{l} = (E_y)_{x+dx}\,dy - (E_y)_x\,dy = [(E_y)_{x+dx} - (E_y)x]\,dy$$

where we are to take the values of E_y at $x + dx$ and x, respectively. The difference between these two values of E_y at the two positions is $(\partial E_y/\partial x)\,dx$, so we can write the line integral of Eq. (12.1) as

$$\frac{\partial E_y}{\partial x}\,dx\,dy = -\mu_0\,\frac{\partial H_z}{\partial t}\,dx\,dy$$

Since this relationship is true for any area $dx\,dy$, we may write

$$\frac{\partial E_y}{\partial x} = -\mu_0 \frac{\partial H_z}{\partial t} \tag{12.19}$$

The next step is to make the converse calculation of the **E** resulting from a varying **H**, through Eq. (12.6). Using Fig. 12.6, we find by a similar calculation,

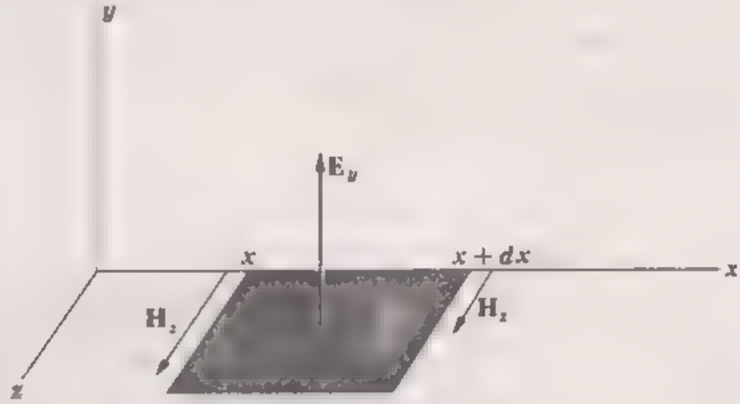


Fig. 12.6 Generation of E_y by a varying H_z.

$$\oint \mathbf{H}\cdot d\mathbf{l} = [(H_z)_x - (H_z)_{x+dx}]\,dz$$

$$= -\frac{\partial H_z}{\partial x}\,dx\,dz = +\epsilon_0 \frac{\partial E_y}{\partial t}\,dx\,dz$$

or

$$\frac{\partial H_z}{\partial x} = -\epsilon_0 \frac{\partial E_y}{\partial t} \tag{12.20}$$

Equations (12.19) and (12.20) relate the space variation of one field to the time variation of the other, and vice versa. When we differentiate the first with respect to x and the second with respect to t, we can combine the information into one equation as follows:

$$\frac{\partial^2 E_y}{\partial x^2} = -\mu_0 \frac{\partial^2 H_z}{\partial x\,\partial t}$$

$$\frac{\partial^2 H_z}{\partial x\,\partial t} = -\epsilon_0 \frac{\partial^2 E_y}{\partial t^2}$$

$$\therefore \frac{\partial^2 E_y}{\partial x^2} = \epsilon_0\mu_0 \frac{\partial^2 E_y}{\partial t^2} \tag{12.21}$$

When we differentiate in the opposite order, we find that the same equation holds for H_z:

$$\frac{\partial^2 H_z}{\partial x^2} = \epsilon_0\mu_0 \frac{\partial^2 H_z}{\partial t^2} \tag{12.22}$$

These last two equations give the result we have been working toward. Comparison with Eq. (12.16) for a transverse wave on a string shows immediately that we have found a differential equation that implies wave propagation. The velocity of propagation along the x direction is given by

$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}} \tag{12.23}$$

The two fields are perpendicular to each other and travel together in the x direction. The agreement between the experimentally observed velocity of light ($\sim 3 \times 10^8$ m sec) and the calculated value based on laboratory measurements of ϵ_0 and μ_0 led Maxwell to propose the electromagnetic nature of light.

Using the results of Sec. 12.4, we write the equation for the electric field component of the electromagnetic wave,

$$E_y = E_0 \sin \frac{2\pi}{\lambda}(x - ct) \tag{12.24}$$

where c is the velocity of travel and λ the wavelength. This is a

Fig. 12.7 Schematic representation of a plane electromagnetic wave.

solution of Eq. (12.21). The magnetic component, which is a solution of Eq. (12.22), is given by

$$H_z = H_0 \sin \frac{2\pi}{\lambda}(x - ct)$$

The program of showing how Maxwell's equations led to electromagnetic waves is now completed. We can use Fig. 12.7 to visualize a plane wave. One more step of interest is to show the relationship between the magnitudes of the electric and magnetic vectors in the wave. We proceed by differentiating Eq. (12.24) for a sinusoidal wave, with respect to t. This gives

$$\frac{\partial E_y}{\partial t} = -E_0 \frac{2\pi}{\lambda} c \cos \frac{2\pi}{\lambda} (x - ct)$$

Substitution in Eq. (12.20) gives

$$\frac{\partial H_z}{\partial x} = \epsilon_0 E_0 \frac{2\pi c}{\lambda} \cos \frac{2\pi}{\lambda} (x - ct)$$

which can be integrated with respect to x to give

$$H_z = c\epsilon_0 E_0 \sin \frac{2\pi}{\lambda} (x - ct) = \left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E_0 \sin \frac{2\pi}{\lambda} (x - ct)$$

We have substituted here the value of c given in Eq. (12.23). This gives the result drawn in Fig. 12.7, that E_y and H_z are in phase, and that

$$\sqrt{\mu_0}\, H_z = \sqrt{\epsilon_0}\, E_y \tag{12.25}$$

If the vacuum is replaced by matter, all these equations must be modified by using μ and ϵ for μ_0 and ϵ_0.

Unpolarized plane waves are made up of a superposition of polarized plane waves in which the directions of **E** and **H** are randomly distributed in the yz plane. The **E** and **H** of each component of these plane waves are perpendicular to each other.

12.6 The Transverse Nature of Plane Waves

We now show that for a plane wave traveling in, say, the x direction, the directions of **D** or **E** and **B** or **H** are limited to the transverse direction (in the yz plane). We do this by taking an elementary cube of dimensions dx, dy, and dz as shown in Fig. 12.8. We apply Eq. (12.2) at some instant to the surface of this cube. This assures that the net flux of **D** through the faces of the cube is zero. Furthermore, since we are dealing with a plane wave, **D** does not vary with y or z, so the net flux through the faces $dx\,dy$ and $dx\,dz$ is zero. This leads us to the conclusion that the net flux through the

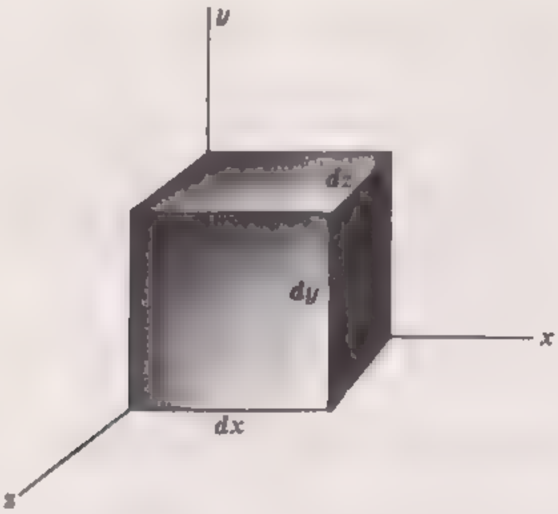


***Fig. 12.8** Elementary cube used in proof that* **D** *and* **B** *are transverse in a plane wave.*

faces $dy\,dz$ is also zero, since the total flux is zero. Now the flux through the left-hand face $dy\,dz$ is $-D_x dy\,dz$ (the negative sign results from the negative sign of the outward normal of this face). The flux through the right-hand $dy\,dz$ is $D_x dy\,dz$. It follows that D_x must have the same value at both faces for all times, and therefore it cannot vary with x. Only the space-varying field enters into the traveling wave, so the wave has no field component in the direction of travel. Any D_x which exists is not involved in the wave. A similar proof regarding B_x is obtained through application of Eq. (12.3). We now have shown that both **D** and **B** (or **E** and **H**) are limited to the yz plane and are therefore transverse to the direction of propagation.

12.7 Propagation of Energy, the Poynting Vector

The propagation of electric and magnetic fields through space implies the transmission of energy, since E and **H** fields involve stored energy. In this section we define and evaluate the energy propagation vector **S**, called the *Poynting vector*. This is the rate of energy flow across a unit area placed perpendicular to the flow direction. In optics this is often called the *intensity*. Its meaning may be understood by considering a volume surrounded by a closed surface and writing an expression for the rate of change of energy within the volume. Let the electromagnetic energy within the volume be U; then the rate of change of this energy is

$$\frac{\partial U}{\partial t} = -\int_{CS} \mathbf{S} \cdot d\mathbf{A} \qquad (12.26)$$

Thus the rate of change of energy within the volume is the surface integral of **S** taken over the surface surrounding the volume. Here the usual symbol for an element of surface area, $d\mathbf{S}$, has been replaced by $d\mathbf{A}$ to avoid confusion with the Poynting vector **S**. The negative sign is used to allow for the fact that the integral on the right-hand side of the equation is negative when energy is entering the volume.

Our problem is to evaluate **S** for an electromagnetic wave. We proceed by considering a volume $dx\,dy\,dz$ (Fig. 12.9) through which

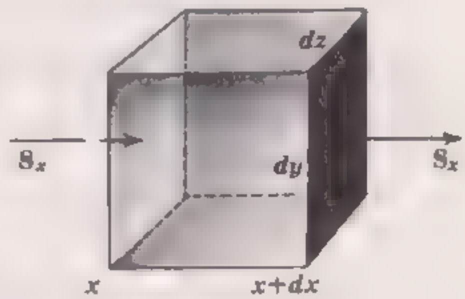


Fig. 12.9 Energy flow of a plane wave through an elementary cube.

a plane wave is passing in the x direction. We express the energy within the volume in terms of **E** and **H** and then relate the rate of change of this energy to the flux of **S** across the surface surrounding the volume. From our previous results, the energy in the volume is

$$U = \left(\frac{\epsilon_0 E_y^2}{2} + \frac{\mu_0 H_z^2}{2}\right) dx\,dy\,dz \tag{12.27}$$

and by differentiating with respect to time, we get

$$\frac{\partial U}{\partial t} = \left(\epsilon_0 E_y \frac{\partial E_y}{\partial t} + \mu_0 H_z \frac{\partial H_z}{\partial t}\right) dx\,dy\,dz$$

We have chosen, as before, a plane wave with **E** and **H** in the y and z directions, respectively. We now replace the time derivatives of E_y and H_z by space derivatives, using Eqs. (12.19) and (12.20). Hence

$$\frac{\partial U}{\partial t} = -\left(E_y \frac{\partial H_z}{\partial x} + H_z \frac{\partial E_y}{\partial x}\right) dx\,dy\,dz$$

$$= -\frac{\partial}{\partial x}(E_y H_z)\,dx\,dy\,dz \tag{12.28}$$

We next put $\int_{CS} \mathbf{S} \cdot d\mathbf{A}$ in the same form as Eq. (12.28) so that $\mathbf{S}$ may be evaluated in terms of E_y and H_z. For a plane wave traveling in the x direction, energy flow must be in the x direction only (since E_y and H_z depend only on x and t), so the flow of energy into and out of the volume is through the $dy\,dz$ faces only, and $\mathbf{S}$ is in the x direction. Thus

$$\begin{aligned}\int_{CS} \mathbf{S} \cdot d\mathbf{A} &= (S_x dy\,dz)_{x+dx} - (S_x dy\,dz)_x \\ &= [(S_x)_{x+dx} - (S_x)_x] dy\,dz\end{aligned}$$

The term in brackets may be written as $(\partial S_x/\partial x)\,dx$, so the last equation may be written

$$\int_{CS} \mathbf{S} \cdot d\mathbf{A} = \frac{\partial S_x}{\partial x}\,dx\,dy\,dz \tag{12.29}$$

We may now compare Eqs. (12.28) and (12.29) through Eq. (12.26), to find

$$\frac{\partial}{\partial x}(E_y H_z)\,dx\,dy\,dz = \frac{\partial S_x}{\partial x}\,dx\,dy\,dz$$

and since this equation holds for any arbitrary volume, it must be that

$$\frac{\partial}{\partial x}(E_y H_z) = \frac{\partial S_x}{\partial x}$$

Integration with respect to x gives

$$S_x = E_y H_z \tag{12.30}$$

plus a constant, which may be neglected.

For the general case, $\mathbf{S}$ is given by the vector product

$$\mathbf{S} = (\mathbf{E} \times \mathbf{H}) \qquad \text{joules/sec-m}^2 = \text{watts/m}^2 \tag{12.31}$$

The Poynting vector $\mathbf{S}$ thus gives the magnitude of the energy flow. The direction of energy flow is perpendicular to the plane containing $\mathbf{E}$ and $\mathbf{H}$ and is in the direction of the vector $(\mathbf{E} \times \mathbf{H})$.

In Eq. (12.31), $\mathbf{E}$ and $\mathbf{H}$ refer to instantaneous values. In the case of sinusoidally varying fields, the average value of $\mathbf{S}$ is given by

$$\mathbf{S} = \mathbf{E}_{eff} \times \mathbf{H}_{eff}$$

or, since $\mathbf{E}_{eff} = \mathbf{E}_0/\sqrt{2}$ and $\mathbf{H}_{eff} = \mathbf{H}_0/\sqrt{2}$, where $\mathbf{E}_0$ and $\mathbf{H}_0$ are amplitudes,

$$\mathbf{S} = \tfrac{1}{2}(\mathbf{E}_0 \times \mathbf{H}_0)$$

In some situations, $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ evaluated over some arbitrary area may give rise to misleading results. Suppose, for instance, that a charged capacitor is placed between the poles of a permanent magnet oriented so that $\mathbf{E} \perp \mathbf{H}$. The Poynting vector $\mathbf{S}$ evaluated over some area perpendicular to $\mathbf{S}$ would then be finite and would seem to indicate a flow of energy. However, the only situation in which the Poynting vector is associated with a measurable quantity is that in which there is a net flow into or out of a volume element. For example, we might evaluate $\int_{CS} \mathbf{S}\cdot d\mathbf{A}$ over a surface surrounding a radio antenna that is radiating electromagnetic waves, in order to determine the total radiated energy. In the case of a capacitor with a fixed charge placed in a static magnetic field, the surface integral evaluated over a closed surface surrounding any volume must always be zero, since no energy is being radiated away.

12.8 *Example, Poynting Vector*

The use of the Poynting vector is valid whenever there is a flow of energy into or out of a volume. We show an example in which $\mathbf{E}$ and $\mathbf{H}$ are constant in time. Consider a cylindrical resistor (Fig. 12.10) of length L, radius r, and resistance R, across which a potential difference V is applied. The rate at which energy is dissipated in this resistor is i^2R. We show that

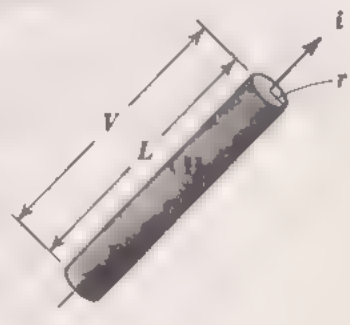


Fig. 12.10 Calculation of energy flow into a resistor. The Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ is directed inward at the surface of the resistor.

$$\int_{CS} \mathbf{S}\cdot d\mathbf{A} = i^2R \tag{12.32}$$

That is, the flow of energy through the surface of the resistor, as measured by the integral of the Poynting vector, accounts for the rate of conversion of electric energy to heat within the resistor.

The calculation involves the determination of **E** and **H** over the surface of the resistor. **E** is pointed along the curved surface and has the value $\dfrac{V}{L} = \dfrac{iR}{L}$. Lines of **H** make circles around the resistor and, from Ampère's circuital law, have the value $H = i/2\pi r$ at the surface. Application of $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ shows that **S** is pointed inward at the curved surface of the resistor (and is parallel to the end surfaces). Evaluation of the integral over the entire resistor surface then gives

$$\int_{CS} \mathbf{S} \cdot d\mathbf{A} = \int EH\, dA = \frac{iR}{L}\frac{i}{2\pi r} 2\pi r L = i^2 R$$

Since **S** is parallel to the end surfaces, the integral over these surfaces is zero. We have proved the statement of Eq. (12.32).

12.9 Generation of Electromagnetic Waves

From the results of the last section it follows that the requirement for generating electromagnetic waves is the simultaneous production of an oscillating **E** and **H** field oriented so as to give a finite value to $\mathbf{S} = \mathbf{E} \times \mathbf{H}$. The simplest arrangement for doing this is the oscillating dipole. This is any arrangement that allows for the periodic linear displacement of charge along some line so as to charge the two ends oppositely and to reverse the charge periodically. It is equivalent to setting up an oscillating sinusoidal current in a conducting wire. We can readily see how this will produce both **E** and **H** fields. It is shown schematically in Fig. 12.11,

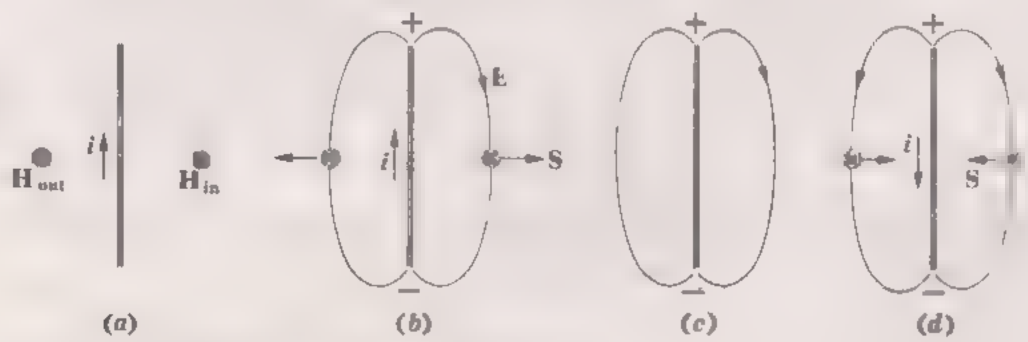


Fig. 12.11 Oscillating dipole, showing fluctuating Poynting vector during half-cycle.

where a wire is sketched at a series of times during half a cycle of alternating current. At (*a*) the wire is neutral, but with an upward current just beginning. The direction of the **H** field due to this current is found by use of the right-hand rule, and the cutting of the plane of the paper by a circular **H** loop around the wire is indicated. At (*b*) the time is advanced and the current is still upward, but now some net charge has collected at the ends of the wire. This produces an electric field, and a line of **E** is sketched to indicate its direction. At the same time the Poynting vector **S** is drawn appropriate to $\mathbf{E} \times \mathbf{H}$, showing that as the charges separate, stored energy is increasing in the space around the wire. The next sketch (*c*) shows the instant of maximum charge separation when the current and hence **H** are zero. Finally, at (*d*) the current has reversed but the charge is not yet reversed in sign. Here the Poynting vector is inward, showing the return of the energy stored in the field. At a first glance it looks as though the fact that **E** and **H** are 90° out of phase would result in the periodic reversal of the direction of the energy flow and no net energy flow outward. This would indeed be true if the field at a point away from the dipole at a given instant depended on the current and charge distribution at the dipole at that instant. However, there is a time lag, depending on the velocity c and on the distance from the dipole, between the setting up of a particular current and charge distribution and the appearance of the consequent fields at a given point. It is this time lag that allows some of the energy in the region around the dipole to continue to travel outward even when conditions of charge and current at the dipole indicate a Poynting vector directed inward to the dipole. This is in many ways remindful of the generation of water waves on a pond by an oscillation at a point on its surface. If the motion at a point in the pond were communicated instantly to the entire pond, there would be no waves.

A rather different way of looking at this problem is to examine the fields of a moving charge under two conditions: first, when it moves with constant velocity, and second, when it is accelerated. We shall find that only when charges are accelerated is there a net radiation of energy. Figure 12.12 shows the two cases. In the sketch of Fig. 12.12*a*, a positive charge has been and is moving at constant velocity. A few of the radial **E** lines are shown and also some of the **H** lines coming out of the plane of the paper. Poynting

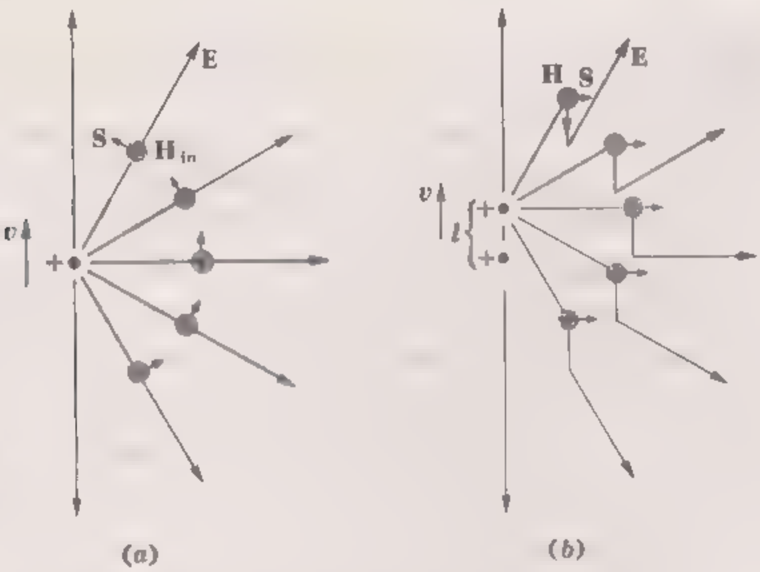


Fig. 12.12 Field and energy flow for a moving charge: (a) Charge moving with a constant velocity; (b) charge moving with a sudden acceleration, returning to constant velocity.

vectors are also drawn showing a continual flow of energy from behind, toward the direction of travel of the moving charge. Detailed calculation shows that this flow of energy is just equal to the flow of energy implied by the advance of the static field of the moving charge, and therefore no radiation is implied. The case of an accelerated particle as in Fig. 12.12*b* is very different. Here we assume that after traveling at constant velocity, the positive charge is suddenly accelerated for a short period, after which it continues with an increased but constant velocity. It is thus a distance l ahead of where it would have been without this sudden change in velocity. This process must put a kink in the lines of **E** as shown, which travel outward. We have shown how the vectors **S** now indicate the generation of a radial flow of electromagnetic energy.

12.10 Reflection and Refraction of Electromagnetic Waves at a Dielectric Boundary

We give one example of the application of the ideas we have developed to an important problem in optics. When a light beam is incident on a dielectric boundary, some of the incident energy is reflected and the remainder is transmitted (refracted) through the second medium. We discuss only the simple case of normal incidence here. Our problem is to determine the relative intensities of the reflected and refracted waves when the incident wave is normal to a plane dielectric boundary. Figure 12.13 shows the geometry of the problem and the quantities involved. The boundary between the two media is the yz plane at $x = 0$. The electric,

magnetic, and Poynting vectors of the incident (E, H, S), reflected (E_1, H_1, S_1), and refracted (E', H', S') plane waves are shown. The regions on either side of the boundary are characterized by the indices of refraction n_1 and n_2, respectively. The index of refraction, usually used in optics to describe the dielectric behavior

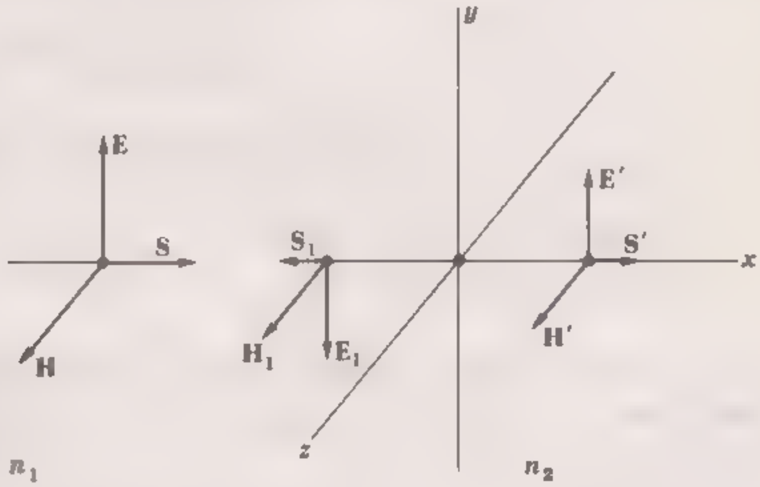


***Fig. 12.13** Reflection and refraction of electromagnetic waves at a dielectric boundary.*

of materials, is defined as the ratio of the velocity of light in free space, c, to the velocity v in a given medium. Thus

$$n = \frac{c}{v} \tag{12.33}$$

This can be related to the electric and magnetic properties of a material as follows. The velocity of light in free space is $c = 1/(\epsilon_0\mu_0)^{1/2}$, and in a medium with permittivity ϵ and magnetic permeability μ, it is $v = 1/(\epsilon\mu)^{1/2}$. The permeability of most optically transparent bodies is very close to μ_0, so we can write

$$n = \left(\frac{\epsilon\mu}{\epsilon_0\mu_0}\right)^{1/2} \simeq \left(\frac{\epsilon\mu_0}{\epsilon_0\mu_0}\right)^{1/2} = \left(\frac{\epsilon}{\epsilon_0}\right)^{1/2} = K^{1/2}$$

where K is the dielectric constant of the material.

In the diagram the $\mathbf{E}_1$ vector, for the reflected wave, is drawn out of phase with the incident $\mathbf{E}$, and $\mathbf{H}_1$ has been made in phase with $\mathbf{H}$. Clearly, either the $\mathbf{E}$ or the $\mathbf{H}$ vector of the reflected wave must be reversed in order that the Poynting vector be in the backward direction. For the moment let us consider our choice arbitrary.

A further requirement that must be satisfied is that the ratio

between the **E** and **H** vectors in each wave must have the value we worked out earlier,

$$\sqrt{\epsilon}\,E = \sqrt{\mu}\,H \qquad \text{or} \qquad H = n\left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E \tag{12.34}$$

From this we have

$$H = n_1\left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E \qquad H' = n_2\left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E' \qquad H_1 = n_1\left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E_1$$

Finally, the boundary conditions must be satisfied:

$$D_{n1} = D_{n2} \qquad B_{n1} = B_{n2}$$

$$E_{t1} = E_{t2} \qquad H_{t1} = H_{t2}$$

For the special case of plane waves normal to the surface, the conditions on D_n and B_n are inoperative since these quantities are both zero.

When we apply the boundary conditions on **E** and **H** and use in addition the relations between **E** and **H** in each wave, we can write

$$E - E_1 = E'$$

$$H + H_1 = H' \rightarrow n_1E + n_1E_1 = n_2E'$$

Solution of these equations for E_1 and E' yields

$$E' = \left(\frac{2n_1}{n_2 + n_1}\right)E \qquad \text{and} \qquad E_1 = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)E$$

When we solve instead for H' and H_1, we get, similarly,

$$H' = \left(\frac{2n_2}{n_2 + n_1}\right)H \qquad \text{and} \qquad H_1 = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)H$$

E_1 and H_1 are both positive if $n_2 > n_1$. This means that the assumption we made in assigning directions to E_1 and H_1 were correct if $n_2 > n_1$ and that these directions will be reversed if $n_2 < n_1$ The flux of energy in, per unit area, is $\mathbf{S} = \mathbf{E} \times \mathbf{H} = EH$. The flux out is readily shown by the addition of $E_1H_1 + E'H'$ to be equal to the same quantity. Thus conservation of energy is satisfied by this result.

For non-normal incidence, the polarization direction needs

to be taken into account in order to determine the divison of energy into reflected and refracted beams. We leave this and other boundary-value problems for study in optics.

12.11 Electromagnetic Waves in Waveguides

The problem of conveying alternating currents from one place to another by using conducting wires becomes increasingly difficult as the frequency increases. There are two sources of difficulty. The first is the skin-depth effect discussed in Chap. 8. Eddy currents tend to confine the a-c current to the outer skin of the conductor, causing the effective resistance to increase and resulting in large power loss along the wire. The second source of trouble lies in the fact that electromagnetic radiation increases as f^4, which results in much of the energy being radiated away instead of being conducted to the place where it is wanted. At frequencies above about 3,000 Mc/sec (at the lower end of the microwave frequency range) these two effects rule out the practical use of simple wire conductors for the transmission of electric energy. One solution to the problem of losses by radiation is to use a coaxial line such as shown in Fig. 12.14. The energy at high frequencies is transmitted as an electro-

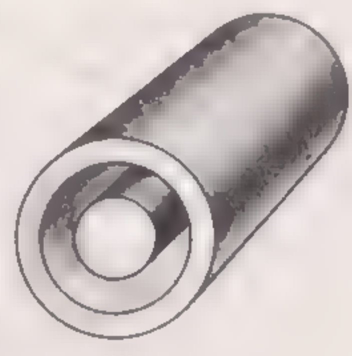


***Fig. 12.14** Coaxial line used for high-frequency transmission.*

magnetic wave passing down the space between the two conductors. Since the radiation cannot penetrate the outer conductor, no energy is lost by radiation. There is one difficulty at very high frequencies, however. It is necessary to support the inner conductor somehow, and this is usually done by filling all or part of the space between conductors with a dielectric material. The difficulty is that at high frequencies most practical dielectric materials are lossy; that is, they absorb part of the energy from the electromagnetic waves.

For the highest microwave frequencies there is an elegant solution to the transmission problem. This is the use of hollow waveguides. These are like coaxial lines except that the central conductor is omitted. In waveguides a suitable radiator sets up an electromagnetic wave that travels within the guide rather like a plane wave in free space. The big difference is that in the guide the wave is altered from a plane wave, as is required in order that the electric and magnetic boundary conditions at the surface of the conductor be satisfied. In general, there are many different modes or shapes of electromagnetic waves that can satisfy the boundary conditions and propagate energy down a guide. It is possible, however, to use a geometry and size of guide that, for a limited range of frequencies, allows only one mode of transmission. We discuss the most commonly used mode, in a rectangular guide such as shown in Fig. 12.15.

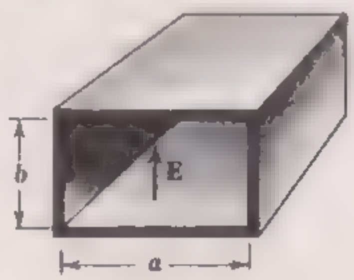


***Fig. 12.15** Hollow waveguide used for microwave transmission.*

It is a straightforward problem in electromagnetic theory to determine the possible electromagnetic waves that can propagate in such a hollow conducting guide. We need only combine Maxwell's equations with the boundary conditions at the metal surface to obtain the possible solutions. We do not attempt this task here but instead describe the results in a frequency range for which only one solution is possible. We shall call this the *dominant* mode in the guide. Let us first discuss the configuration of the oscillating electric field of the wave. This field points in the direction of the smaller dimension of the guide, b. This means that the wave is polarized. Over any cross section of the guide, E is a maximum along the center line of the guide and falls off in intensity to zero next to the narrow wall. This fall-off to zero is required to satisfy the condition that E at a surface is zero. (We are making the assumption here that the walls are perfect conductors.) This is shown in Fig. 12.16, where electric field vectors (not lines of force) are drawn. E varies sinusoidally down the guide, as shown in the side view. The electric field configuration is thus very similar to that in a polarized plane

Fig. 12.16 Electric field vectors in waveguide.

wave. The two modifications are that $E_{\parallel}$ goes to zero and that the region of field is limited to the cross section of the guide. The modification in the magnetic induction field is more severe. It is still everywhere perpendicular to E, as in a plane wave, but because of the boundary condition the B lines form closed loops in planes parallel to the broad face of the guide.

The boundary condition on the lines of varying B is that $B_{\perp}$ at the surface of a conductor goes to zero. This follows at once from the condition that $E_{\parallel}$ goes to zero. We argue that B must be perpendicular to E (from Maxwell's equations, as we showed in the discussion of a plane wave), so if $E_{\parallel}$ goes to zero, so must $B_{\perp}$. Thus the effect of the conducting boundary is to modify the infinite plane wave, in which lines of B form closed loops only at infinity, in such a way that the B loops are closed within the guide. Lines of B and E of a traveling wave are shown in Fig. 12.17. Use of the

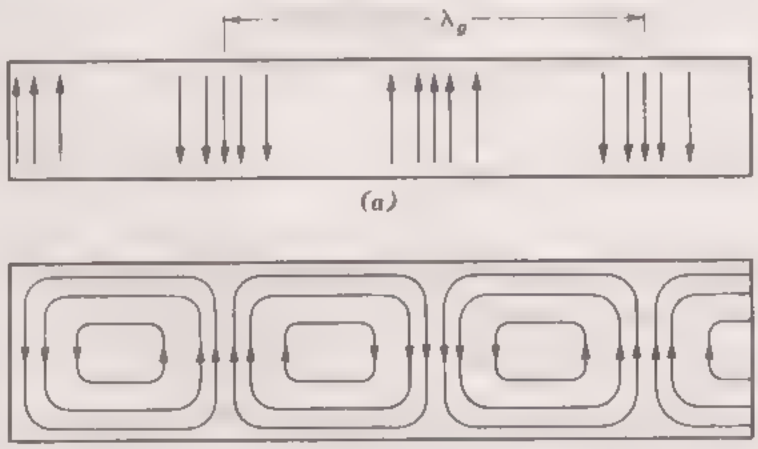


Fig. 12.17 Lines of E and B in waveguide: (a) Lines of E, side view; (b) lines of B, top view.

Poynting vector $\mathbf{S} = (1/\mu_0)(\mathbf{E} \times \mathbf{B})$ shows that this wave is traveling to the right.

Two important features of electromagnetic waves in guides are that the velocity with which energy travels along the guide is different from c, the velocity of a wave in free space, and that for each kind of mode there is a definite cutoff frequency below which waves cannot propagate through the guide. If we measure the characteristic of an electromagnetic wave in free space, we find that $\lambda\nu = c$, where c is the velocity of light. If, however, we measure

the wavelength in a waveguide, λ_g, we find that for a given frequency ν, λ_g is greater than is obtained in free space. (This measurement can be made by setting up standing waves by reflecting the wave at the end of the guide.) For the wave mode involved here, λ_g is related to λ in free space by the equation

$$\lambda_g = \frac{\lambda}{[1 - (\lambda/2a)^2]^{1/2}} \tag{12.35}$$

where a is the broad dimension of the guide.

At first glance we might be worried that $\lambda_g > \lambda$ seems to imply a velocity in the guide greater than the velocity of light in free space, in direct contradiction to relativity theory. However, the relativistic limitation on velocities refers to mass or energy transport. In a waveguide or in any situation where the velocity depends on frequency, we must consider two kinds of velocity. The first is the *phase velocity* v_p, obtained by measuring the distance between peaks in the wave. This gives us the guide wavelength λ_g. The other is the velocity with which energy propagates, which we could obtain by sending a short pulse of energy down the guide and measuring the time for the pulse to go a certain distance. This is called the group velocity v_{gr}. In free space $v_p = v_{gr}$, but, in general,

$$v_p v_{gr} = c^2$$

Since the phase velocity in a guide is greater than c, it follows that the group velocity is less than c. From Eq. (12.35) we see that when the free-space wavelength λ increases to $2a$, the guide wavelength becomes infinite and v_{gr} goes to zero. That is, energy cannot be transmitted down the guide for frequencies lower than that corresponding to $\lambda = 2a$. This limiting value is called the *cutoff frequency*.

PROBLEMS

12.1 Starting with Eqs. (12.19) and (12.20), which relate time and space variations of **E** and **H**, show that

$$\frac{\partial^2 H_z}{\partial x^2} = \epsilon_0\mu_0 \frac{\partial^2 H_z}{\partial t^2}$$

12.2 Show that the differential equation of Prob. 12.1 can be satisfied, for example, by a sine wave in H_z traveling in the negative x direction,

if the velocity is given by

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$

12.3 An a-c generator is connected to a parallel-plate capacitor made of circular disks of area A. As a result, the charge q on the plates is $q = q_0 \sin \omega t$. The lines of **H** induced by the resulting displacement current are circles with centers on the axis of symmetry of the capacitor. Show that the magnetic field intensity at any point between the plates is given by

$$H = \frac{q_0 r \omega}{2A} \cos \omega t$$

where r is the distance from the axis of symmetry. Neglect edge effects.

12.4 Find the frequencies of electromagnetic waves having the following wavelengths in free space:
a 10^3 m (long-wave radio)
b 1 m (short-wave radio)
c 3 cm (microwaves)
d 10^{-4} m (infrared)
e 5,000 A (optical) (1 angstrom = 10^{-10} m)
f 0.1 A (X rays)
g 10^{-3} A (gamma rays)

12.5 A plane radio wave travels in the x direction and is plane-polarized with its electric vector in the y direction. Its frequency is 1 Mc/sec. The average power propagated by the wave is 20 watts/m^2.
a Find the wavelength of the wave.
b Find the amplitudes of **E** and **H** for this wave.

12.6 Show that if we describe a traveling wave [Eq. (12.24)] by means of its *wave number* $k = 2\pi/\lambda$ and its angular frequency $\omega = 2\pi f$, the wave is described by the equation

$$E_y = E_0 \sin (kx - \omega t)$$

12.7 Sunlight strikes the earth, outside its atmosphere, with an intensity of 2.0 cal/cm^2-min. Calculate the peak values of **E** and **B** for sunlight at the earth.

12.8 Find the lowest frequency for which an electromagnetic wave can be transmitted through a rectangular waveguide whose inner dimension perpendicular to the E field is 1 cm.

12.9 A rectangular guide has a width (perpendicular to the **E** field) of 1.0 cm. For the dominant mode [for which Eq. (12.35) holds], what is the guide wavelength for an electromagnetic wave for which the free-space wavelength is 1.25 cm? What is the phase velocity of this wave in the guide? What is the group velocity with which energy propagates in the guide?

12.10 Show that for frequencies at which vacuum tubes are ordinarily operated, the vacuum displacement current is negligible compared with the current carried by electrons. Take, for example, a tube with cathode and plate area each 1 cm^2, spaced 5 mm apart, passing 10 ma (rms) at a frequency of 10^6 cps. Assume parallel-plate geometry and assume that the cathode is at fixed potential and the rms a-c voltage across the tube is 100 volts. What is the magnitude of the rms displacement current through the tube?

THIRTEEN

Conduction of Electricity in Gases and Magnetohydrodynamics

13.1 Introduction

In this chapter we discuss the application of the laws of electricity and magnetism to two important fields of physics. We first consider the conduction of electricity in gases. Whenever ions and electrons are present in a gas, currents can flow as a result of an externally applied electric field. When fields are high enough, electrons may gain enough energy to ionize atoms or molecules in the gas as a result of collisions, and a number of processes may occur that provide a continuous source of ions and electrons. Depending on the conditions, there may result a self-maintaining *glow* or *corona discharge* or a *spark* breakdown in the gas. We discuss the atomic processes occurring in these phenomena and give a brief résumé of the conditions under which each type of discharge occurs.

The second application is to the field of magnetohydrodynamics. There can be a strong interaction between the dynamics of a highly conducting fluid (such as an ionized gas containing equal numbers of positive and negative charges) and magnetic fields.

This interaction can affect the mass motion of the fluid, and it gives rise to a variety of phenomena of interest in the fields of astrophysics and geophysics. It is also of great importance in the problem of harnessing thermonuclear fusion energy. We begin by a study of the motion of individual charges in a magnetic field, extending the discussion of Sec. 6.7. We shall see how the recently discovered Van Allen belt of trapped charges in the ionosphere can be explained on the basis of magnetic forces. Following this, we study the role of magnetic forces in conducting fluids and see how they give rise to magnetohydrodynamic waves in the fluid. Finally, we shall investigate briefly the way in which fluids can be confined by means of magnetic fields, a problem of major importance in the development of thermonuclear energy sources.

13.2 Low Field Processes

The passage of electricity through gases depends on the presence of charge carriers (+ and − ions and electrons). Thus in low applied electric fields, the current flowing between electrodes inserted in a volume of gas depends on ions and electrons normally present in the gas as the result of the ionization caused by cosmic rays or other sources of high-energy radiation. Currents can be enhanced by providing an artificial source of ionization such as X rays. The drift motion of the charge carriers is interrupted by *elastic* collisions with atoms or molecules of the gas. By elastic collisions we mean those which conserve energy and momentum

The drift motion of carriers in an applied electric field is conveniently described by the *mobility* μ of the carrier, where μ is defined by the equation

$$v_d = \mu E \tag{13.1}$$

where v_d is the average velocity in the direction of the field E. The mobility depends on the effective cross section for collisions between the carriers and gas atoms and on the charge and mass of the carriers and gas atoms. The mobility of carriers can be discussed in a way very similar to our presentation of the resistivity of a metal in Sec. 7.3. We found for electrons in a metal that the drift velocity is given by

$$v_d = \frac{e}{m}\tau E \tag{7.7a}$$

where τ is the mean relaxation time. By comparison of Eqs. (13.1) and (7.7*a*), we find for the mobility

$$\mu = \frac{\epsilon}{m}\tau \tag{13.2}$$

The mobility of electrons and ions in a gas depends on the relaxation time, which in turn depends on *collision cross sections* and on gas pressure. By collision cross section we measure how close one particle must approach another in order that its path be appreciably deflected. For example, the collision cross section between a billiard ball and a small particle (of size negligible compared with the billiard ball) is the cross-section area of the billiard ball. Electric forces between particles modify this simple result. The cross section for collisions between electrons and atoms varies markedly for different kinds of atoms and also depends strongly on electron velocity. The mobility of gas ions also varies with the kinds of atoms involved.

Ion mobilities can be measured directly by studying current flow in a gas under low applied fields. One method involves the production of ions in the gas close to one electrode as shown in Fig. 13.1. A square-wave alternating voltage is applied between

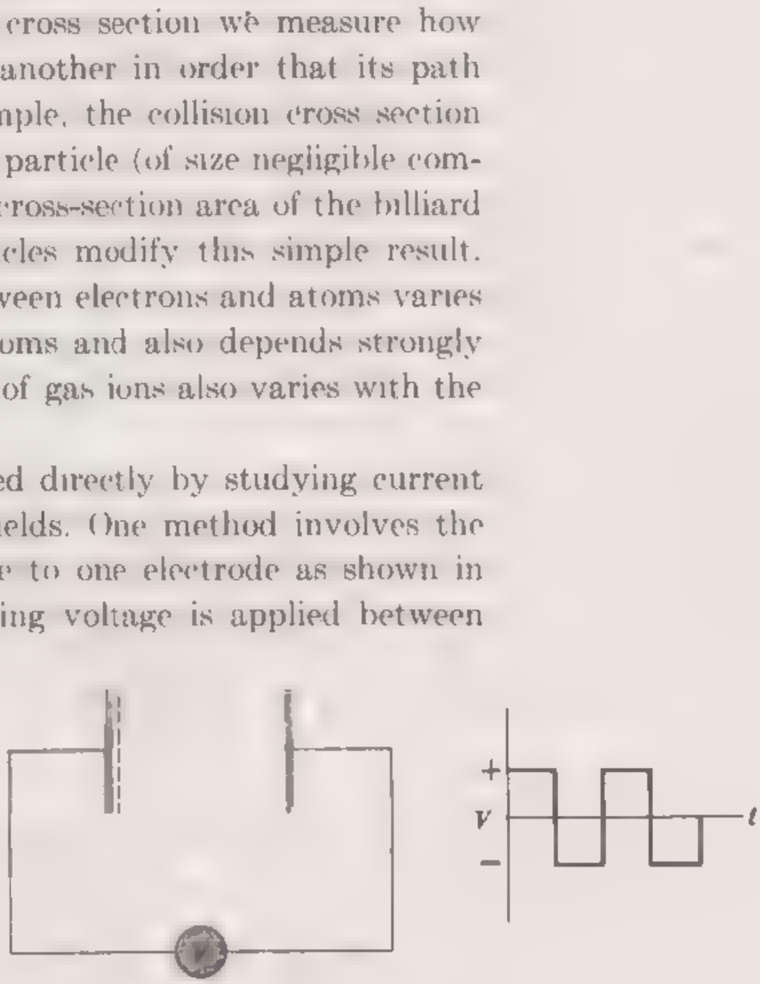


***Fig. 13.1** Apparatus for measuring the mobility of ions in a gas. Ions are produced near one electrode by an external source of ionization. Square-wave applied voltage is shown.*

two electrodes. For low applied voltages, ions produced by X rays near one electrode cannot reach the opposite electrode before the field is reversed and hence are not collected by the electrode. As the voltage of the square wave is gradually increased, the drift velocity finally reaches a value that allows the ions to reach the collecting electrode. Calculations of E and knowledge of the time during which the voltage is applied by the square wave allow the drift velocity and the mobility to be calculated.

Gas mobilities range from a few tenths of a centimeter per second per volt per centimeter for heavy molecular ions up to

about ten centimeters per second per volt per centimeter for H_2^+ at atmospheric pressure and room temperature. Since at constant temperature the time between collisions is inversely proportional to gas pressure, mobility varies inversely with pressure. Both negative and positive ions are produced by ionizing radiation in a gas. The negative ions result from electron attachment to neutral molecules. There are a number of processes by which positive and negative ions and electrons recombine to form neutral atoms or molecules. Since the rate of recombination depends on the density of charge carriers, the recombination process is exponential in character.

When a constant source of ionization (say an X-ray beam) is present, a current between plane electrodes placed in a gas results from the collection of the ions produced when a steady field is applied between the electrodes. The current depends on the rate of production of ions by the source as modified by recombination in the gas and diffusion of some ions out of the field region between the plates. With low applied fields the current is proportional to the voltage applied, but with increased voltage the rate of increase of current with voltage becomes less (Fig. 13.2). This saturation effect results from the fixed rate of production of ions. Once all the ions produced are being collected, the current can no longer increase until secondary processes, as discussed in the next section, begin to play a role.

13.3 High Field Processes, Ionization by Collision

When the voltage applied between parallel plates in a gas provided with a source of ionization, as discussed above, is increased beyond the values shown in Fig. 13.2, the current-voltage curve rises steeply, and further increase in voltage leads to a spark or catastrophic current increase, called *electrical breakdown*. The onset of increased current marks the beginning of *ionization by collision*.

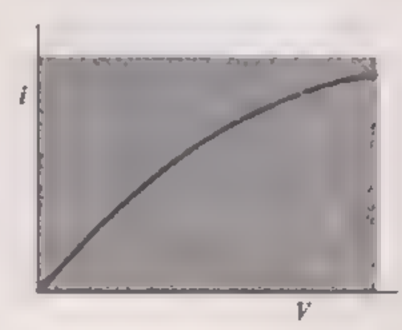


Fig. 13.2 Saturation of current through a gas in low fields. Ions and electrons that carry the current are produced by an external source.

The voltage at which this occurs depends on the electrode shape and spacing and on the kind of gas and its pressure. Conditions for the onset of such processes are arrived at when the value of E/p, where E is the applied field and p is the gas pressure, reaches a critical value that depends on the kind of gas. A number of kinds of processes can be involved, but they all depend on the fact that charge carriers in fields above the critical value gain enough energy between collisions to initiate some secondary process that adds to the number of available ions in the gas.

We can easily see why the critical condition depends on E/p. Thus, suppose at some pressure the critical applied field has been reached, so that ions are sufficiently accelerated between collisions that their kinetic energy is enough to initiate some secondary ion-producing process. When the pressure is doubled, reducing the mean free path between collisions by a factor of 2, twice the field must be applied if the ions are to gain as much energy as they did at the lower pressure. Thus, whatever the secondary mechanism, the critical field depends on the ratio E/p. In air at atmospheric pressure, the critical field is about 15,000 volts/cm, or $E/p = 20$, where the field is in volts per centimeter and p is in millimeters of mercury. Approximate critical values of E/p for other gases are given in Table 13.1.

Table 13.1 Values of E/p in Volts/cm/mm Pressure Leading to Ionization by Collision

Gas	E/p
Air	**20**
H_2	**10**
A	**5**
Ne	**2**

We now discuss the principal mechanisms of secondary ion production. The most important secondary process, and the first to appear as the applied field is increased, is ionization by electron collision. At the critical value of E/p some of the most energetic electrons gain enough kinetic energy in the applied field to cause ionization when they collide with an atom or molecule. That is, an

electron is knocked off the atom, leaving a positive ion and an extra electron. If events are fortuitous, both electrons will make ionizing collisions during their next free paths, producing two more electrons and two more positive ions. This kind of process can continue, giving rise to an avalanche of electrons initiated by a single electron. The form of this kind of increase is easily derived. Suppose one electron makes α electrons by collision in traveling 1 cm in the direction of the force due to the external field. The increase dn in the number of electrons in a distance dx caused by n electrons will then be

$$dn = n\alpha\, dx \tag{13.3}$$

When we write this as $dn/n = \alpha\, dx$ and integrate, we find $\ln n = \alpha x + \ln \text{const}$ If n_0 is the number of electrons we start with, when $x = 0$, we can evaluate the constant as n_0 and write the result in the form

$$\frac{n}{n_0} = e^{\alpha x} \tag{13.4}$$

Since the current is proportional to the number of electrons present, we may also write

$$i = i_0 e^{\alpha x} \tag{13.5}$$

α is called the *first Townsend coefficient* and gives the increase in current caused by ionization. α itself is not constant but varies with E/p.

Two other processes that can occur at higher values of E/p are now considered. The first is the liberation of photoelectrons at the cathode (negative electrode). We discuss the photoelectric effect in detail in Chap. 14. Here it is sufficient to point out that electromagnetic waves (light) of high enough energy are able to give enough energy to an electron in a metal surface to enable it to escape. In the collision process it may be that an electron does not ionize an atom but gives it enough energy to leave it in an *excited state*. A very energetic electron may not only ionize an atom but also leave it in an excited state. An excited state could be one in which an orbiting electron is removed from its normal position into another orbit of higher energy. When in due course, and generally very rapidly, this electron drops back into its normal or ground state, the excess energy is given off in electromagnetic

energy. If this energy falls on the cathode, an electron can be emitted that is then able to form its own avalanche of electrons as it travels through the gas. Electrons are also emitted at the anode, but they are pulled back to the anode by the applied field.

Another secondary process results from collisions of positive ions with the cathode. Some of the kinetic energy of the positive ion can be transferred to an electron in the metal surface, allowing it in some cases to excape into the gas. As in the former process, this electron can then produce others by avalanche formation.

We show in Appendix E the nature of the current flowing when both primary multiplication and secondary processes occur. If n_0 is the rate at which electrons are produced near the cathode by an external source of ionization, n the total number of electrons per second reaching the anode, and γ the probability that a secondary electron will be freed at the cathode by a secondary process associated with each primary multiplication event, we find

$$n = n_0 \frac{e^{\alpha x}}{1 - \gamma e^{\alpha x}} \tag{E.2}$$

or

$$i = i_0 \frac{e^{\alpha x}}{1 - \gamma e^{\alpha x}} \tag{E.3}$$

This equation gives the important criterion for self-maintaining conductivity in a gas. As the denominator approaches zero, the current i increases, and finally a situation is reached in which current flows regardless of the initial current i_0. The criterion for *voltage breakdown*, or self-maintaining current, is then just

$$1 - \gamma e^{\alpha x} = 0 \qquad \text{or} \qquad \gamma e^{\alpha x} = 1 \tag{13.6}$$

This means that, starting with one electron at the cathode, enough ion pairs are formed in the multiplication process so that secondary processes produce at least one new electron at the cathode. The multiplication process is thus self-maintaining. Depending on the type of discharge, this can lead to either a glow discharge, a corona discharge, an arc, or a catastrophic increase in current, or spark, reducing the resistance of the gap essentially to zero.

13.4 The Glow Discharge

The glow discharge occurs in gases at less than atmospheric pressure, from say 0.1 mm Hg up to perhaps 100 mm pressure. This kind of process is characterized by a visible glow from the gas in the region of the cathode and a less striking illumination of the gas between electrodes. The primary multiplication takes place near the cathode, and secondary electrons are produced at the cathode by positive ion collisions. Details of the glow discharge are rather complicated, owing to the distortion of the uniform electric field by space-charge effects. In the main body of the gas, positive and negative charges are about equal, and fields are low enough so that little multiplication takes place. Emission of ultraviolet light is characteristic of the glow discharge. Fluorescent lights are glow-discharge tubes in which the ultraviolet light is converted to visible light by a fluorescent coating on the tube surrounding the gas volume.

13.5 The Corona Discharge

In the corona discharge, one electrode is sharp. This produces a concentrated high electric field in the region of the point, in which the active corona process originates. If the point is positive, electrons pulled into the very high field region near the point ionize many atoms, and a high-density space charge results from the slowly moving heavy positive ions left behind by the more rapidly moving electrons. Thus serious local distortions of the original field can occur. In addition, the region of intense ionization emits intense ultraviolet radiation as a result of excitation processes occurring along with the ionization process. Consequently, secondary ionization can occur not only at the opposite electrode but also within the gas volume. As a result of this secondary ionization and of the large space-charge effects, *streamers* can propagate out into the region of the gap where ionization could not normally occur in the original unmodified field. These streamers are luminescent linear paths of heavy ionization, resulting from both primary and secondary multiplication processes. If conditions are such that the streamers can propagate completely across the gap, they can provide a highly conducting path that leads to a complete voltage breakdown across the gap, and a high-current spark or arc can result.

In a negative point corona, electrons quickly gain energy as they leave the point after being released by positive-ion bombardment or by ultraviolet radiation. In the high field region there is heavy ionization both by the primary collision process and by secondary ionization in the gas. Under some conditions, visible streamers are also formed in the region near the point. The processes in these point corona discharges are very complicated; they often involve intermittent currents as a result of the relatively slow motion of the positive-ion space charge, which can very greatly alter the electric field distribution.

13.6 The Arc Discharge

In contrast with glow or corona discharge, in the arc discharge the production of secondary electrons at the cathode is usually a thermal process. Large currents of positive ions lose kinetic energy upon striking the cathode, and as a result the cathode temperature can rise to the point where thermionic emission of electrons is possible. Arc discharges are characterized by very large currents, often of many amperes, and relatively low voltage between electrodes. They can occur at atmospheric pressure.

13.7 Particle Motion in a Magnetic Field

We now turn to the discussion of some aspects of magnetohydrodynamics. Before taking up the study of the motion of an ionized fluid, or *plasma*, in a magnetic field, we discuss the motion of individual ions, adding to the simpler cases discussed in Chap. 6.[1] We assume that the ions are in a very-low-pressure gas, so that the disturbing effects of collisions on ion motion can be neglected.

The force on a charged particle in a magnetic field was found to be

$$\mathbf{F} = e(\mathbf{v} \times \mathbf{B}) \tag{6.25}$$

where e is its charge and $\mathbf{v}$ its velocity. This force gives rise to circular motion if $\mathbf{v}$ has a component perpendicular to the magnetic

[1] This discussion follows fairly closely that in Lyman Spitzer, Jr., "Physics of Fully Ionized Gases," Interscience Publishers, Inc., New York, 1959.

induction field **B**. The angular velocity in this motion is

$$\omega_c = \frac{eB}{m} \tag{6.26}$$

where ω_c is called the cyclotron (angular) frequency. In general, where the velocity has components both parallel and perpendicular to the magnetic field, we can write

$$\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp}$$

giving the velocity in terms of two components. The velocity $v_{\parallel}$ is not affected by the magnetic field, according to Eq. (6.25), so this component is constant. The radius of curvature of the perpendicular component is given by $r = v_{\perp}/\omega_c$ or

$$r = \frac{mv_{\perp}}{eB}$$

so the radius depends only on the perpendicular component of the velocity. Combining these two motions, we see that in general the motion of a charged particle in a uniform field is a helix with constant pitch, with the axis of the helix pointed in the field direction.

We now discuss the motion under a variety of perturbing conditions. Since we have seen that the components of motion along and perpendicular to the magnetic induction field can be separated, we do not consider the simple uniform motion parallel to the field in the following discussions. To begin with, we consider the situation of a uniform electric field **E** perpendicular to a uniform magnetic induction field **B**. Let **v** be the velocity of the particle perpendicular to **B**. We first write the equation for the total force acting on the charged particle,

$$F = e[\mathbf{E} + (\mathbf{v} \times \mathbf{B})] \tag{13.7}$$

This combines the magnetic force of Eq. (6.25) with the electric force given by Eq. (2.3). Using Newton's law, we may rewrite this as

$$m\frac{dv}{dt} = e[\mathbf{E} + (\mathbf{v} \times \mathbf{B})] \tag{13.8}$$

The easiest way to solve this equation is to make the substitution

$$\mathbf{v} = \mathbf{v}' + \frac{\mathbf{E} \times \mathbf{B}}{B^2} \tag{13.9}$$

We do this because we shall find that the motion perpendicular to **B** can be described in terms of a circular motion **v**′ and a uniform translational velocity $(\mathbf{E} \times \mathbf{B})/B^2$. The procedure is a scheme that works only because we are able to guess the proper substitution formula, but the fact that it works successfully fully justifies its use. We proceed by substituting $v' + (\mathbf{E} \times \mathbf{B})/B^2$ for v in Eq. (13.8) and find

$$m\frac{dv'}{dt} = e\left[\mathbf{E} + (\mathbf{v}' \times \mathbf{B}) + \frac{1}{B^2}(\mathbf{E} \times \mathbf{B}) \times \mathbf{B}\right] \tag{13.10}$$

The meaning of $(\mathbf{E} \times \mathbf{B}) \times \mathbf{B}$ can be understood by examining Fig. 13.3. When **E** is along the x direction and **B** is along the y

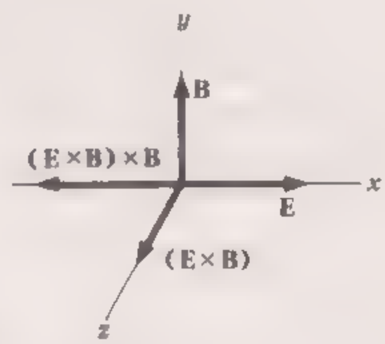


Fig. 13.3 Diagram illustrating the direction of the vector $(\mathbf{E} \times \mathbf{B}) \times \mathbf{B}$.

direction, the vector $(\mathbf{E} \times \mathbf{B})$ is in the z direction. Taking the cross product of this vector with **B**, we find the resultant vector points in the $-x$ direction. Thus $(\mathbf{E} \times \mathbf{B}) \times \mathbf{B} = -B^2\mathbf{E}$ or

$$\frac{1}{B^2}(\mathbf{E} \times \mathbf{B}) \times \mathbf{B} = -\mathbf{E}$$

This substitution then reduces Eq. (13.10) to

$$m\frac{dv'}{dt} = e(\mathbf{v}' \times \mathbf{B}) \tag{13.11}$$

Since this has the same form as Eq. (6.25) for the motion of a charge in a magnetic field only, we see that **v**′ describes simple circular motion. However, the total velocity is given by this circular motion plus a uniform drift velocity perpendicular to both **E** and **B** (in the z direction in Fig. 13.3). The magnitude of the drift velocity $(\mathbf{E} \times \mathbf{B})/B^2$ is E/B. The nature of the total motion is shown in

Fig. 13.4 Effect of an electric field on the circular motion of charged particles in a uniform magnetic field. A drift velocity is added to the circular motion as shown.

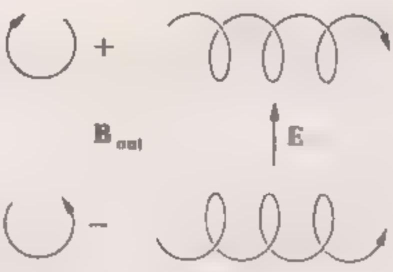


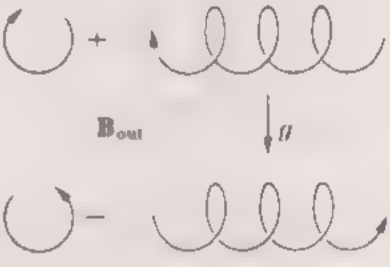
Fig. 13.4. It is notable that an electric field applied in a given direction produces a drift in a direction perpendicular to **E**. In the more general case where **E** has a component parallel to **B**, this component adds an acceleration to the component of velocity parallel to **B**.

A second case of interest is that of a gravitational field perpendicular to the uniform **B** field. This contrasts with the previous electric-field case in that the extra force is changed from the sign dependent $e\mathbf{E}$ to $mg_\perp$, where $mg_\perp$ is the gravitational force perpendicular to **B**. When we make this change in the equation for the drift velocity v_D, we find

$$v_D = \frac{mg_\perp}{eB} = \frac{g_\perp}{\omega_c} \tag{13.12}$$

The drift velocity is in the direction of $\mathbf{g}_\perp \times \mathbf{B}$ for a positive particle and is reversed for a negative particle. Figure 13.5 shows the nature of the motion.

Fig. 13.5 Effect of a gravitational field on the circular motion of charged particles in a uniform magnetic field. Contrast with electric-field case results from the independence of charge sign of the gravitational force.



A final case of interest is that of the motion of charges in an inhomogeneous field. Figure 13.6 shows the kind of motion that

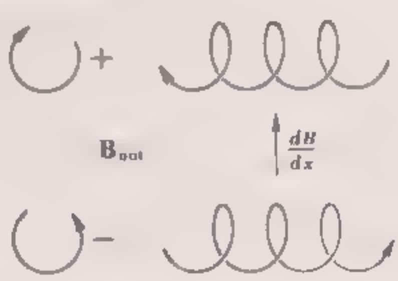


Fig. 13.6 Drift motion is also caused by variation in the magnetic field intensity in a direction perpendicular to the field.

results when **B** varies linearly in the x direction perpendicular to the direction of **B** The qualitative nature of this motion is easily understood on the basis that the curvature is greater in the region of higher magnetic field and less in the lower field region. As a consequence, the circular motion in a uniform field is distorted as shown. We do not attempt a quantitative discussion of this case.

Our next task is to discuss the interesting case of particle motion in a region in which the lines of force are converging That is, **B** is increasing as we move along the direction of lines of force. Before examining this situation, however, we consider an important theorem regarding circular motion of a charged particle in a magnetic field. This theorem states that if **B** varies slowly in space and time, the magnetic moment resulting from particle motion in the field is constant. Let the magnetic moment resulting from circular charge motion in the field be μ. Then

$$\mu = iA = e\frac{\omega_c}{2\pi}\pi r^2 = \frac{\frac{1}{2}mv_\perp^{\,2}}{B} \tag{13.13}$$

We prove the theorem first for the case of **B** uniform in space but varying with time. The time variation of flux within the area of the circular path gives rise to an induced emf $\mathcal{E}$ according to Faraday's law, and we can write

$$\mathcal{E} = \oint \mathbf{E}\cdot d\mathbf{l} = -\int \frac{dB}{dt}\,dA \tag{12.1}$$

The rate at which work is done on the moving charge is

$$\frac{dW}{dt} = \frac{d}{dt}\left(\frac{1}{2}mv_\perp^{\,2}\right) = \mathcal{E}i = \frac{e\omega_c}{2\pi}\pi r^2\frac{dB}{dt} \tag{13.14}$$

We have evaluated $\mathcal{E}$ from Eq. (12.1) and used $i = e\omega_c/2\pi$ for the current due to the rotating charge. Comparison with Eq. (13.13) shows that

$$\frac{dW}{dt} = \frac{d}{dt}\left(\frac{1}{2}mv_\perp^{\,2}\right) = \mu\frac{dB}{dt} \tag{13.15}$$

When we multiply Eq. (13.13) by B, we get $\mu B = \frac{1}{2}mv_\perp^{\,2}$. Differentiation of this with respect to time gives

$$\frac{d}{dt}(\mu B) = \mu\frac{dB}{dt} + B\frac{d\mu}{dt} = \frac{d}{dt}\left(\frac{1}{2}mv_\perp^{\,2}\right)$$

But in view of the result of Eq (13.15), the term $B\, d\mu/dt$ must equal zero, so μ is a constant, and we have proved the point we were after.

The case of **B** constant in time but increasing as we move along lines of **B** is now considered. Figure 13.7 shows the

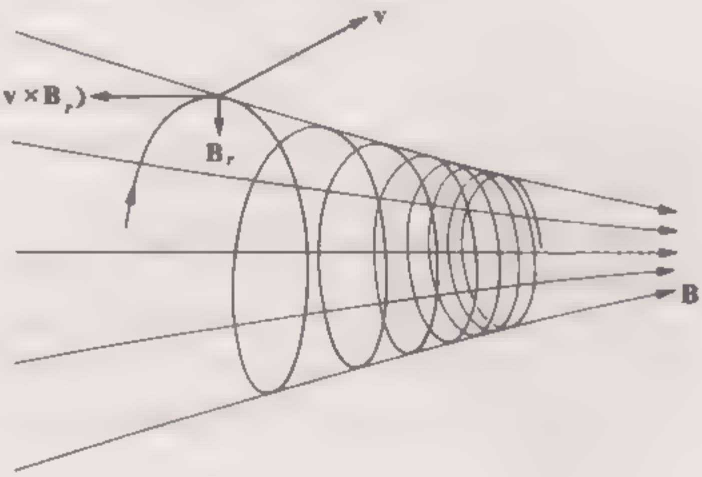


Fig. 13.7 Helical path of a charged particle in a converging magnetic field. A force (v × B$_r$) acts to slow down the drift motion along the field direction.

helical path of a charged particle in a converging magnetic field, and we discuss the nature of the motion in this situation. The proof that, in this case also, the magnetic moment of the circular motion is constant will not be given. It is true, however, as long as B changes are small over distances comparable to the radius of circular motion, and we use the result in the following discussion.

Whenever there is a radial component B_r pointing toward the guiding center of the helical path, there is a force acting to decelerate the motion along the axis of the spiral, given by $(\mathbf{v} \times \mathbf{B}_r)$. This tends to retard the drift velocity in the field direction. We show that the motion in the field direction not only is decreased but actually goes to zero and then reverses. Converging lines of **B** can thus reflect the motion (whence the term *magnetic mirror*). However, since the total magnetic force is always perpendicular to the velocity of the particle, no work is done on the particle and therefore its velocity stays constant in magnitude. Thus, if $\mathbf{v}$ is the total velocity of the particle and $\mathbf{v}_D$ is its velocity component along the axis of the helix and $\mathbf{v}_\perp$ is the velocity perpendicular to the axis, we have

$$\mathbf{v} = \text{const} = \mathbf{v}_\perp + \mathbf{v}_D \qquad (13.16)$$

Thus as $\mathbf{v}_D$ decreases, $\mathbf{v}_\perp$ must increase so as to keep $\mathbf{v}$ constant in magnitude. From Eq. (13.13) for the magnetic moment of the

circular motion, we see also that if μ is to remain constant, $v_\perp^2/B$ must stay constant so that the fractional increase of $v_\perp^2$ is proportional to the fractional increase in **B**. If we call θ the angle made by the helical motion with respect to the axis of the helix, or the angle

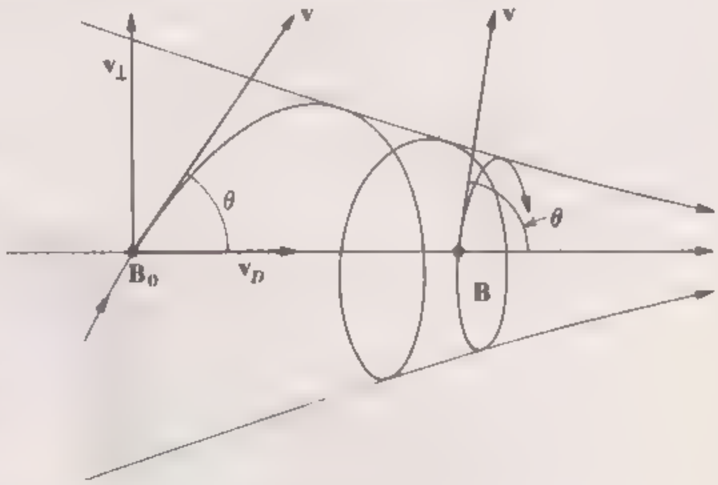


Fig. 13.8 Helical path in a converging field showing change in angle of advance from θ_0 to θ when field has changed from B_0 to B.

by which the particle advances along the helix (Fig. 13.8), we can write

$$\frac{v_\perp}{v} = \sin\theta \qquad \text{or} \qquad \frac{v_\perp^2}{v^2} = \sin^2\theta \tag{13.17}$$

The value of θ changes as the particle moves along the helix. When, at some point where the field is $\mathbf{B}_0$, the value of θ is θ_0, and for some point further along, the field is **B** and the angle is θ, we can compare the angles of advance in the two locations, using Eqs. (13.16) and (13.17), which give

$$\frac{\sin^2\theta_0}{B_0} = \frac{\sin^2\theta}{B} \qquad \text{or} \qquad \sin^2\theta = \frac{B}{B_0}\sin^2\theta_0 \tag{13.18}$$

When the particle has moved far enough along the helix so that $\dfrac{B}{B_0} = \dfrac{1}{\sin^2\theta_0}$, $\sin^2\theta = 1$. This means that $v_\perp = v$, so all the motion is perpendicular to the axis of the helix, and the drift velocity has become zero. However, there is still a backward force, so the drift reverses after stopping and we find that the particle motion has been reflected by the converging lines of **B**. The kinetic energy of the particle remains constant during this reflection process.

This reflection of particle motion by converging lines of magnetic field is of importance in the explanation of the recently dis-

covered high density of charged particles called the Van Allen belt, which surrounds the earth out to distances four or five times the radius of the earth. Information on this region of high radiation intensity was obtained in 1958 from the flights of the satellites Explorer I, III, and IV and Sputnik III and Mechta.[1] The existence of this high radiation intensity is explained by the trapping of high-energy protons and electrons of cosmic origin by the geomagnetic field. Figure 13.9 is a sketch of the form of the region, which con-

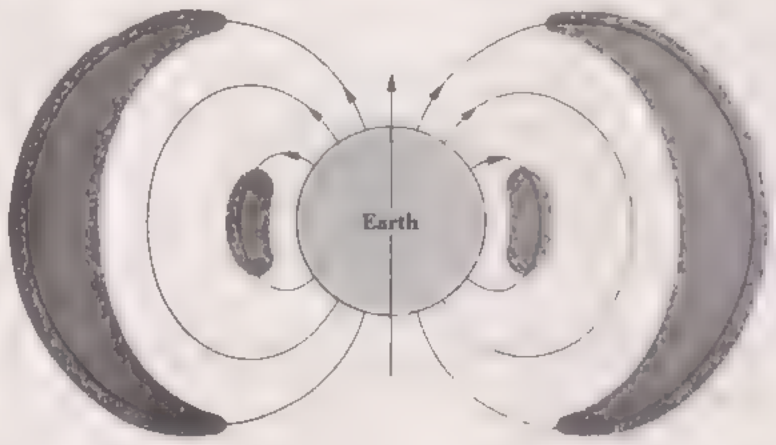


***Fig. 13.9** The Van Allen belt of charged particles captured by the earth's magnetic field. The inner torus contains protons, and the outer one contains electrons.*

sists of an inner region of high proton density and a more distant region containing electrons. These regions encircle the earth, and they tend to follow the curvature of the lines of magnetic field around the earth. The trapping of the charged particles is easily understood on the basis of the classical mirror effect we have considered.[2] Figure 13.10 is a sketch (not to scale) of the way in which the converging lines of magnetic field out from the earth produce reflections of the helical motion of charged particles and hence effectively trap them. The theory of this effect has long been understood and was available immediately when the experimental data were obtained from satellite observations.

Another application of the mirror effect was postulated by Fermi and Alfvén to account for the cosmic acceleration of cosmic-ray particles. Fermi suggested that in interstellar clouds the magnetic field could be greater than in the intervening regions. If

[1] Information on this work and further references can be found in J. A. Van Allen, The Geomagnetically Trapped Corpuscular Radiation, *J. Geophys. Research*, 64:1683–1689 (1959).

[2] An extensive treatment of this and other problems in magnetohydrodynamics is given in H. Alfvén, "Cosmic Electrodynamics," Oxford University Press, New York, 1950.

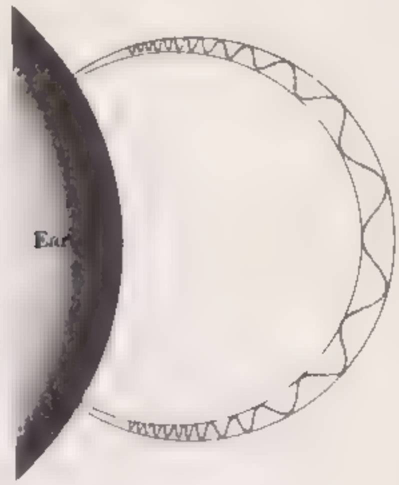


Fig. 13.10 Helical path of charged particle captured in earth's magnetic field. Particle spirals back and forth between the two magnetic mirrors formed by converging field lines. Drawing is not to scale.

that were so, then particles in the region between two clouds could be trapped by the converging fields, as we have seen. If the two clouds are moving toward each other, it can be shown that the particles gain kinetic energy each time they are reflected by the magnetic mirror. This mechanism would explain at least one of the causes of very-high-energy charged particles in cosmic rays.

13.8 Magnetohydrodynamic Waves

We now turn from the discussion of single-particle motion in a magnetic field to a consideration of some of the characteristics of the motion of a fluid of such high ion density that it is highly conducting. We shall consider the case of a *plasma* in which the density of positive and negative particles is equal. If a magnetic field is present, moving charges (hydrodynamic motion) give rise to induced E fields, which in turn give rise to current. These currents are acted upon by the magnetic field, and the resultant forces change the state of motion. This connection between mass motion and electromagnetic fields can set up *magnetohydrodynamic* waves.

We can see this by means of a very crude model. Suppose a conducting fluid is immersed in a uniform magnetic field. Suppose to begin with that all the fluid is at rest except for a column moving with constant velocity v in the negative direction, into the paper (Fig. 13.11). Consider this column to be of infinite extent along the z direction. We wish to show how electric and magnetic forces resulting from the motion of this column cause consequent motion of other parts of the fluid, and in fact set up waves of motion that propagate through the medium.

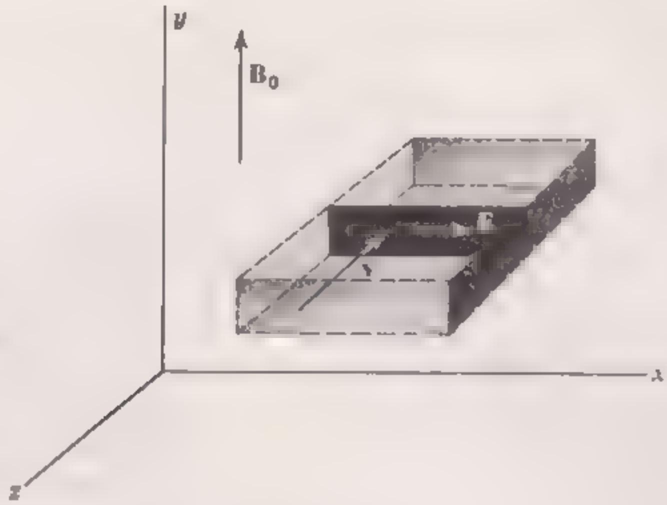


***Fig. 13.11** A column of highly conducting fluid moving perpendicular to a magnetic field, showing induced electric field E in moving column.*

We have seen in Sec. 8.2 that when a conductor moves with a velocity v perpendicular to a magnetic field **B**, magnetic forces tend to separate charges on the two ends of the conductor. These forces, given by $F = qvB$, continue to separate charges until an electric field E is set up in the conductor, which acts in opposition to the magnetic forces. In equilibrium, then, there is an electric field set up in the conductor given by

$$E = -\frac{F}{q} = -vB$$

The exact statement of relative directions is

$$\mathbf{E} = -(\mathbf{v} \times \mathbf{B}) \tag{13.19}$$

As a result of this field, currents are set up in the moving conducting fluid as well as in the adjacent stationary fluid, somewhat as sketched in Fig. 13.12. Though we have considered equal

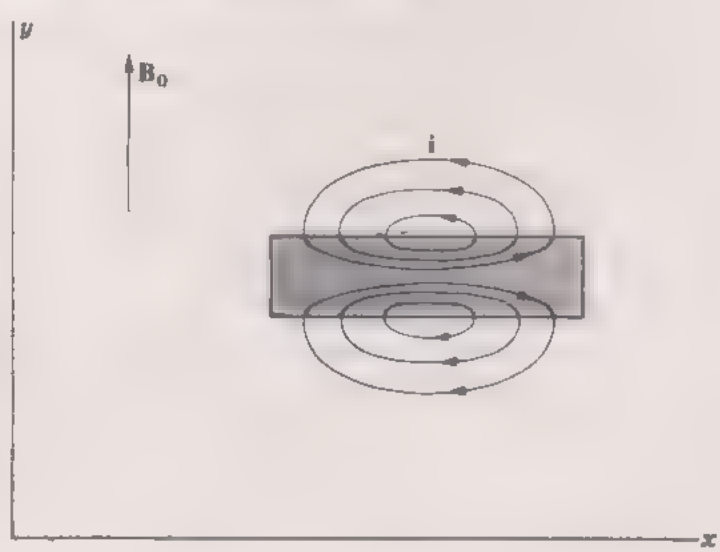


***Fig. 13.12** Current resulting from induced electric field sets up forces that retard moving column and accelerate adjacent regions. This action sets up magnetohydrodynamic waves.*

densities of positive and negative carriers, they move in opposite directions to give currents in the same direction. Even though the net charge is zero, both signs of charge contribute to the current.

These currents will interact with the magnetic field, and as a result there are forces acting on the charges of the fluid. The force on each current element is

$$\mathbf{F} = i\,d\mathbf{l} \times \mathbf{B} \tag{6.3}$$

Since the current inside the moving column of fluid is in the positive x direction, Eq. (6.3) gives a force in the positive z direction, out of the paper, which opposes the original motion in the negative z direction. In the adjacent static columns of fluid, above and below the moving column in the diagram, the current is in the opposite direction, so the force is in the negative z direction. Thus the original column tends to be slowed down, and adjacent columns tend to gain velocity in the original direction of motion. The events described show how this kind of electromagnetic coupling provides a mechanism by which motion in one region is transferred to surrounding regions. It is this action which is responsible for magnetohydrodynamic waves.

13.9 Flow of a Conducting Medium in a Magnetic Field

We now discuss the interaction between the flow of a conducting fluid and a magnetic field. Motion of the fluid perpendicular to the lines of magnetic field is seriously impeded by the circular motion of the charges, while motion parallel to the field is unaffected. In a highly conducting medium, in fact, we can assume that the magnetic lines are "frozen" in the fluid and move along with any transverse motion of the fluid. In effect, a magnetic field tends to *contain* a plasma and prevent it from diffusing outward.

A great deal of work is now being done to investigate this situation in connection with attempts to produce controlled nuclear fusion energy. The problem is to raise the temperature of a plasma, which will undergo fusion with a consequent release of energy, to a value high enough ($\sim 10^8$ °K) to allow the process to occur. The hope is that by means of a magnetic field, the plasma can be confined to a volume away from the surface of the container,

so as to prevent the loss of heat that occurs if the plasma is in contact with the walls. The difficulty is that many possible field configurations turn out to be unstable and thus are ineffective in containing the plasma. It remains to be seen whether or not there is a practical solution to this problem.

FOURTEEN

Electric and Magnetic Quantum Effects

14.1 Introduction

In this chapter we depart from our consideration of classical electromagnetism, to give a very brief discussion of the impact of *quantum mechanics* on classical electricity and magnetism. We find that there are situations in which electromagnetic radiation is best described in terms of *particles* rather than by waves. Historically, it was with great difficulty that physicists attempted to reconcile the particle nature of radiation with its wavelike properties. Nowadays it is realized that there is no real contradiction in these two aspects, though it is to be expected that different kinds of experiments bring out different aspects. In the following discussion we give some very brief descriptions of certain experiments. Further studies are more appropriate to another course, in which a framework of quantum theory can be built up and the problem can be examined in more detail.

14.2 The Photoelectric Effect

The discovery of the photoelectric effect just before the turn of the century gave one of the most direct cases in which classical theory failed to give a sensible explanation of the observed phenomena. The effect has to do with the ejection of electrons from a metal surface when electromagnetic radiation (light) is incident on the surface. Einstein's explanation of the effect in 1905 was based on the quantum theory of Planck, first published in 1900. On the basis of what had gone before, there was no indication that the very surprising results of the experiments would occur. There were at the time, however, quite a number of experiments pointing in the same direction as the photoelectric effect.

At the time of this development the idea of a work function was well understood. It was clear that to remove an electron from a metal, enough energy has to be given it to allow it to escape over the potential barrier that ordinarily keeps it in the metal. There are a number of known ways of providing this energy. One method of considerable importance is to heat the metal to such an extent that the thermal kinetic energy of the conduction electrons allows some of the most energetic electrons to escape from the surface. This is called *thermionic emission* and is the usual source of electrons in vacuum tubes at present. Another method is called *secondary emission*. When the surface of a metal is bombarded by atoms, ions, or electrons, there are conditions under which a conduction electron can be given enough kinetic energy to allow it to escape. Finally, if a very high negative potential is applied to a metal, the electric field produced at the metal surface may be sufficient to pull the electron over the potential barrier and give what is called *field emission*. This requires fields of the order of 10^6 volts/cm at the surface.

It might be expected that if an electromagnetic wave were incident on a metal surface, there would be conditions under which the oscillatory **E** field of the wave could impart enough energy to an electron to allow it to escape. This is indeed the photoelectric effect, but the details are surprisingly different from those which can be predicted by any classical theory. The first observations of this effect were made in 1887 by Hertz, who found that the sparking in air between two highly charged electrodes occurred at lower voltage when the electrodes were illuminated. However, these experiments were not carried far, since at that time electrons were not yet known

and the background of knowledge was not far enough advanced to allow the experiments to be carried to a useful conclusion.

An experiment like the later one of Lenard in 1902 brings out the contrast between the expected and actual behavior of electromagnetic waves. Figure 14.1 shows two electrodes placed

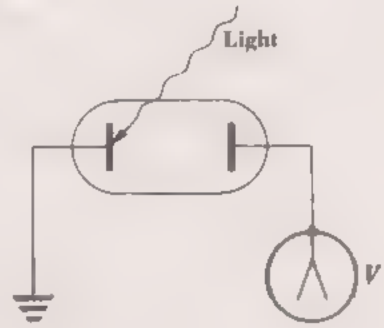


Fig. 14.1 Apparatus used in photoelectric-effect experiment.

in a glass envelope that can be evacuated. One electrode is grounded and the other is connected to an electroscope that measures the potential of the second electrode with respect to ground. When the first electrode is illuminated, energy is imparted to some electrons, which allows them to escape from the metal with some remaining kinetic energy $W = \frac{1}{2}mv^2$. Those electrons which happen to be traveling toward the second electrode travel through the vacuum and are collected on it. This process continues until the charge collected builds up a potential that prevents the electrons from arriving. Figure 14.2 shows a time plot of the way in

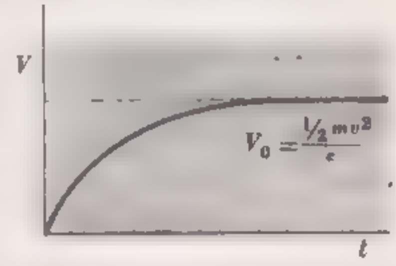


Fig. 14.2 Buildup of charge to equilibrium value in photoelectric-effect experiment.

which equilibrium is reached. This occurs when the initial energy of the emitted electrons, $W = \frac{1}{2}mv^2$, just equals the potential energy eV, which they must gain to reach the second electrode. Thus,

$$V_0 = \frac{\frac{1}{2}mv^2}{e} \tag{14.1}$$

where V_0 is the equilibrium potential of the electrode. [Actually, the applied potential necessary to stop the electrons is modified

by the work functions of the electrodes. Thus if the potential as read on a voltmeter connected between the two plates is called V_s, the effective potential difference is $V_0 = V_s + (\Phi_2 - \Phi_1)/e$, where Φ_1 and Φ_2 are the work functions of the emitter and collector plates, expressed in electron volts.] On a classical basis we expect that if the brightness of the illumination is increased, giving a larger magnitude to the electric field vector in the electromagnetic wave, the electrons will get more initial energy and V_0 will be increased. The fact is that the only result of increasing the illumination intensity is that the equilibrium value is reached more rapidly. That is, more electrons are emitted in a given time, but their energy is unchanged. This result is perhaps only a little less surprising than the further discovery that the only factor that produces a change in V_0 is the frequency f of the incident light. Figure 14.3 shows a plot of the potential V_0 related to the frequency

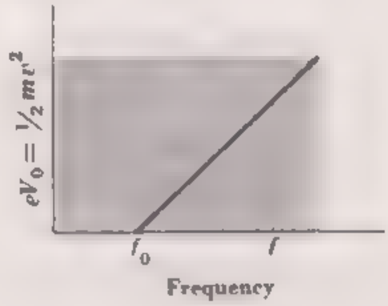


***Fig. 14.3** Electron energy versus frequency of illuminating light in photoelectric experiment.*

of the illumination on a given metal. This result means that the kinetic energy given the electrons is proportional to the difference in frequency from some minimum frequency f_0.

f_0 is found to vary from metal to metal, but the slope of the curve is the same for all metals. At frequencies below f_0 no electrons are emitted. Thus not only is $\frac{1}{2}mv^2 \propto (f - f_0)$ but the proportionality constant is independent of the metal used. We may write

$$\tfrac{1}{2}mv^2 = h(f - f_0) \tag{14.2}$$

where h is the constant of proportionality. This is found to be Planck's quantum constant ($h = 6.62 \times 10^{-34}$ joule-sec). When we let $hf_0 = W_0$, we may write this relationship as

$$hf = \tfrac{1}{2}mv^2 + W_0 \tag{14.3}$$

This may be interpreted as follows: An amount of energy hf is given up by the radiation, of which one part W_0 depends on the kind

of metal; this is used in getting the electron out of the metal while the rest is converted into kinetic energy of the liberated electron. The energy W_0 is the work function for the metal. We are thus entitled to think of the energy in the electromagnetic waves as consisting of discrete *quanta*, each having an energy

$$W = hf \tag{14.4}$$

The discrete or particle-like nature of the energy content is forced on us if we are to understand how the electromagnetic wave is able to impart its energy to a *single* electron in the metal rather than to share it among many. The experiment indicates that the radiation energy either is not imparted to the electron or gives up the entire quantity hf to a single electron. It is usual to speak of these quanta of radiation as *photons*. In so doing, we are emphazing their particle-like properties.

14.3 Atomic Radiation

The result of the photoelectric effect—that electromagnetic energy is in quanta of size hf—is in complete agreement with the known facts of atomic radiation and fits in completely with the atomic model of quantized energy levels. Thus when an electron falls from one allowed energy level in an atom to a lower vacant level, the energy released is radiated electromagnetically as a photon and satisfies the relation

$$\Delta W = hf \tag{14.5}$$

where ΔW is the energy difference between the two states. It is the application of this idea which, in the study of spectroscopy, has allowed the untangling of the great mass of data giving the frequencies of radiation from atoms and has given us the basis for the modern model of the atom.

14.4 X rays and γ rays

In 1895 a very energetic kind of radiation was discovered by Roentgen and was found to penetrate matter that is opaque to ordinary light. These *X rays* are produced when a high-velocity beam of electrons impinges on a metal electrode. A tube for the

production of X rays is shown in Fig. 14.4. Potential differences of from 30 up to 100 kv or more are applied between the electron source and the metal anode. The tube is evacuated so that the electrons reach high energies before they collide with the anode. X rays emerging through the glass walls of the tube were at first

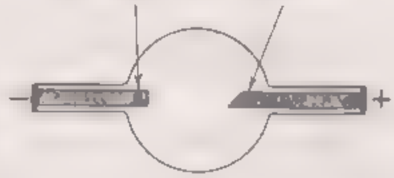


Fig. 14.4 X-ray tube.

studied by their ability to ionize gases and to produce exposure on a photographic plate, and their absorption by various materials was investigated.

One process which occurs is that a high-energy electron can knock an electron from a deep-energy level in an atom of the metal completely out of the atom. When another electron falls back into this level, the energy released by this sudden decrease in potential energy is emitted as a quantum of electromagnetic energy, and an X ray is the result. Another process producing X rays is the sudden stopping of the high-energy electrons by impact with the metal. This sudden deceleration can result in the direct transfer of the kinetic energy of the electron into an electromagnetic quantum or X ray.

In 1913, von Laue reasoned that if X rays were electromagnetic waves, they should exhibit interference effects typical of waves, as observed in the diffraction of light through a grating. He therefore sent a beam of X rays through a salt crystal, in which the periodic spacing between atoms allowed it to act as a diffraction grating. He found a diffraction pattern on a photographic film placed in the beam beyond the crystal that not only showed the expected pattern due to wave interference but also made possible a measurement of the wavelength of the X rays. The necessary knowledge of the spacing between atoms in the crystal (the lattice constant) he obtained from the density of the crystal and Avogadro's number. Whereas the wavelength range of visible light lies between the limits of about 4,000 and 7,000 A (one angstrom is equal to 10^{-10} meter), X rays lie in the wavelength range between about .01 and 100 A. Use of the relation $\lambda f = c$, where λ is the wavelength, f the

frequency, and c the velocity of travel of the waves, gives frequencies between approximately 10^{20} and 10^{16} cps for X rays.

The use of diffraction patterns of single crystals not only has provided an accurate tool for measuring X-ray wavelengths but also has given very detailed information about the spacing and arrangement of atoms in solids.

Another process that releases high-energy quanta occurs in nuclei of atoms. One of the processes by which energy is given off by unstable nuclei is the emission of γ rays. These are like X rays except that they result from nuclear events and are often of still higher energy.

14.5 The Compton Effect

There is a kind of interaction between X rays and matter that exhibits very clearly the particle-like nature of electromagnetic radiation. This interaction was first observed by A. H. Compton in 1923. When X rays are incident on a thin foil of metal, some of the energy of the X ray can be given to an electron and cause it to be ejected from the metal. It can happen that not all the energy is transferred to the electron, and as a result, the X ray proceeds with less than its original energy (and therefore has a lower frequency). The special feature here is that if we treat the X ray as a particle, with energy hf and momentum $p = hf/c$, then the event is completely described as a particle collision process between a photon and an electron. That is, both momentum and energy are conserved, and if direction and energy of the emitted electron are known, the direction and frequency of the scattered X rays can be calculated. The relatively small energy used to free the electron from the metal is insignificant compared with the very much greater energy of the incident X ray, and so may be neglected in making the energy and momentum balance.

14.6 Electron Waves

Present knowledge of atomic structure indicates that in the atom and in solids, it is useful to attribute wavelike properties to particles. The experiments of Davisson and Germer and of G. P. Thomson in 1927 have demonstrated this for electrons in a very revealing way. In these experiments, a beam of electrons is incident on a thin

film of matter, and the pattern of electrons that comes through the film is found to exhibit the typical diffraction pattern of waves, exactly like the diffraction patterns of X rays. Another case in which the wavelike properties of electrons are made apparent is the *electron microscope*. Since electrons can be deflected by electric or magnetic fields, it is possible to focus a beam of electrons by means of appropriately shaped electric fields and magnetic coils, much as a lens focuses light. Such electric or magnetic lenses are used in the electron microscope to produce highly magnified images of thin films through which a high-energy electron beam passes. The fact that the wavelength of sufficiently high-energy electrons is less than the wavelength of light allows higher resolution; also, the wavelike properties of the electrons allow the internal structure of the matter to be inferred from the diffraction patterns observed. Various kinds of crystal imperfections in thin sheets of metals are being studied by this method. Particles as small as viruses, which are much too small to be resolved in ordinary light, are resolved in electron microscopes.

14.7 Magnetic Quantization, the Stern-Gerlach Experiment

During the period of growth of the quantum theory, there were a series of surprises for physicists, which have perhaps not been equaled before or since. We have given short discussions of a few of these and now turn to another, which relates to atomic magnetism. We have already described the sources of magnetism in atoms, the orbital motion of electrons, electron spin, and nuclear magnetism. The experiments of Stern and Gerlach in 1921 were set up to measure the magnetic moments of individual atoms. Prior to this, magnetic measurements had been confined to susceptibility measurements on bulk matter. The principle of the experiment is to measure the deflection of a beam of neutral atoms due to a nonhomogeneous magnetic field through which they pass. Figure 14.5 shows a schematic drawing of the experimental arrangement

This experiment can be used to measure the magnetic moments of atoms. A highly collimated beam of atoms is emitted from the oven source at one end of the apparatus. The atoms travel through an evacuated chamber in a divergent magnetic field produced by appropriately shaped poles of an electromagnet. After passing

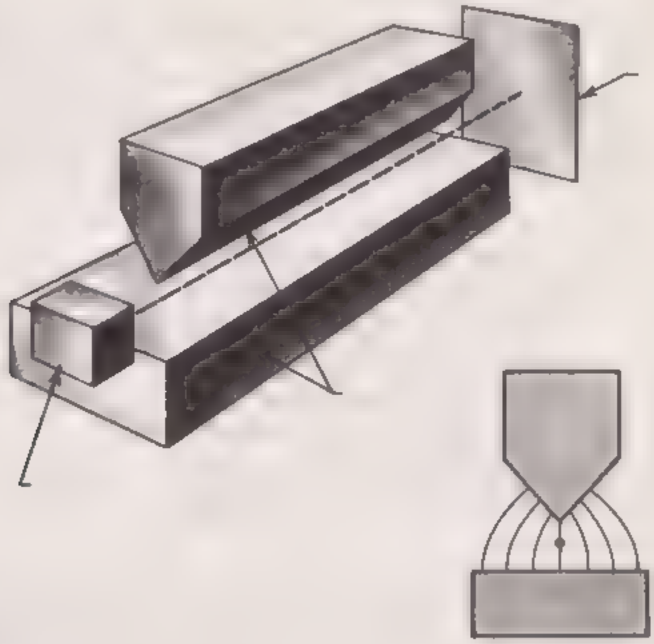


Fig. 14.5 Atomic-beam apparatus.

through the field, the beam hits a fluorescent screen or a photographic plate, which records its position. There is a net force on a magnetic dipole in a nonuniform magnetic field. It is convenient to describe this effect using the polar model of a magnetic dipole (the equivalent of an electric dipole in a divergent electric field). Such a dipole is shown in Fig. 14.6, using q_m for the effective pole

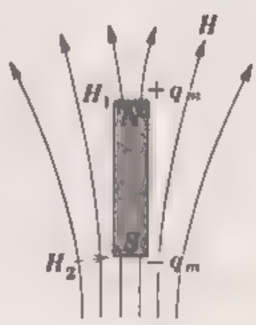
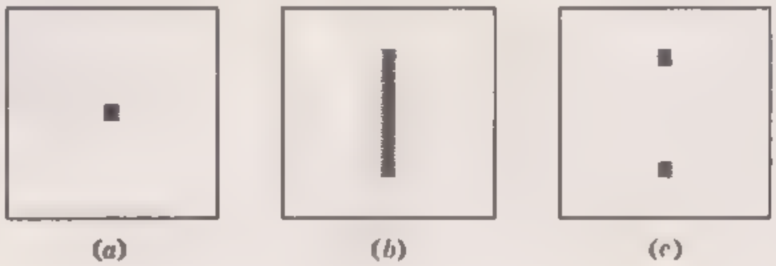


Fig. 14.6 Force on a magnetic moment in a divergent field.

strength (where q_m is defined as in Sec. 9.13) and l for the pole separation. The upward components of the field at the two poles are H_1 and H_2, respectively. Now since the force on each pole is $q_m\mu_0 H_1$ and $-q_m\mu_0 H_2$, the net force is $q_m\mu_0(H_1 - H_2)$, in this case in the downward direction. With the dipole oriented in the opposite direction, the force would be upward. For other orientations, the net force is less since there is less difference between H_1 and H_2 if the dipole is not aligned along the direction of maximum field gradient. The effect of the divergent field is thus expected to be a lengthening of the spot made on the screen by the beam in the

absence of the field. Some atoms would be deflected upward and some downward, to a greater or lesser extent, depending on their orientation in space when they travel through the magnetic field. Figure 14.7 shows three patterns on the screen, the first with no

Fig. 14.7 Atomic-beam traces: (a) Beam in absence of magnetic field; (b) classical expectation for beam after passing through divergent H field; (c) experimental result for beam after passing through divergent H field.



field, the second the expected pattern with the field turned on, and third, the actual pattern observed with, say, lithium atoms. Instead of all possible orientations, only two occur, one parallel to the field and one antiparallel.

This result is another of the phenomena explained by quantum theory. It results from the quantization of the possible orientations of a magnetic dipole in a magnetic field. Another description that amounts to the same thing is to say that the energy levels allowed for the dipole in a field are quantized. Since it takes work to turn a dipole from the direction in which it is aligned with the field into the antiparallel direction, its orientation is expressible in terms of an energy, and it is this energy which is quantized. In the example shown, only two orientations are allowed, although in other situations there may be more. This same technique has been applied to magnetic nuclei by Stern and Rabi and is at present much used in the exact determination of nuclear magnetic moments.

The quantum-mechanical behavior of magnetic dipoles of electrons and nuclei is the basis of an active field of research in solid-state physics at the present time. Samples are placed in a magnetic field where radiation in the high-frequency or microwave region is present. Transitions of the magnetic dipoles from one magnetic state to another are induced by the radiation and result in energy loss from the radiation field to the sample. Results of these experiments give information about the dipoles themselves and about the matter in which they occur. These are called *magnetic resonance* experiments.

FIFTEEN

Units of Measurement

15.1 Introduction

In this book we use the meter-kilogram-second (mks) system of units, which is rather well known and has certain convenient features. However, since much of the important literature in physics and in engineering has been written in *electrostatic* and *electromagnetic cgs* units, it is essential for the student to have some knowledge of these systems. Furthermore, a considerable amount of current writing is in the cgs system, so there is a continuing need for an understanding of both systems.

The important features of the mks system as we have used it are as follows:

1. All experiments, whether purely electric, purely magnetic, or a combination, use the same units. In contrast with this, in the cgs system two subsystems are developed, requiring conversion from one to the other in some situations.

2. In most cases the quantities used in the mks system are identical with the *practical* quantities of electricity and magnetism,

such as amperes, volts, ohms, and farads, which have been for a long time the most common units in practical measurements in both physics and engineering. In the cgs system, the absolute quantities derived from the basic equations must usually be converted to practical units.

3. We have used the *rationalized* mks system, which brings the factor 4π into the fundamental equations. Its presence there gives a somewhat simpler form to many of the frequently used derived equations.

4. The mks system uses meters and newtons instead of centimeters and dynes as in the cgs systems.

In this chapter we examine the cgs system and contrast it with the mks system, showing how to convert from one to the other system.

15.2 The Absolute Electrostatic System of Units (ESU)

The basic force equation is written as

$$F = \frac{q_1 q_2}{r^2} \tag{15.1}$$

in esu, this equation being used to define the absolute unit of charge in esu. When two equal charges q_1 and q_2 are separated by 1 cm in a vacuum, if the force between them is 1 dyne, the charges are each 1 electrostatic unit (esu). Another name for this charge unit is the *statcoulomb*.

In the mks system, the equation is written

$$F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2} \tag{1.5}$$

and the factor $1/4\pi\epsilon_0 = k$ is chosen as 9×10^9 newton-m^2/coulomb^2. This factor makes the charge unit used in electrostatics, the coulomb, agree with the unit of charge as obtained from the magnetic force between current elements. The coulomb is determined from magnetic-force experiments by defining the unit of current as one ampere, equal to one coulomb per second.

We compare the esu or statcoulomb with the coulomb. Confusion is avoided by thinking of two equal charges separated by a

given distance and then using the two equations to evaluate the charge quantity. Thus in the mks system we write

$$q^2\ (\text{coulombs}^2) = \frac{F\ (\text{newtons})\ r^2\ (\text{meters}^2)}{k} \tag{15.2}$$

In the esu system we have

$$q^2\ (\text{esu}^2) = F\ (\text{dynes})\ r^2\ (\text{cm}^2) \tag{15.3}$$

In order to compare these two statements of the same physical situation, we think of the changes required in Eq. (15.2) if we measure F and r in dynes and centimeters. Since 10^5 dynes is equal to one newton, the numerical value of F measured in dynes must be divided by 10^5 in order to get the correct value of coulombs. Similarly, r^2 measured in square centimeters must be divided by 10^4 to give the correct value. Thus we may rewrite Eq. (15.2) as

$$q^2\ (\text{coulombs}^2) = \frac{F\ (\text{dynes})\ r^2\ (\text{cm}^2)}{10^5 \times 10^4 \times 9 \times 10^9} \tag{15.4}$$

Substituting from Eq. (15.3) and taking the square root, we get for Eq. (15.4)

$$q\ (\text{coulombs}) \times 3 \times 10^9 = q\ (\text{esu}) \tag{15.5}$$

This means that an esu unit of charge is very small compared with a coulomb since a given charge is measured as many more esu than coulombs. An alternative expression having the same meaning as Eq. (15.5) is

$$1\ \text{coulomb} = 3 \times 10^9\ \text{esu} \tag{15.6}$$

Equation (15.5) tells how to convert a given answer in coulombs to the same quantity expressed in esu, while Eq. (15.6) gives the ratio between the two kinds of units.

Other quantities derived from the fundamental force equation have units that follow logically from it. We need merely to replace ϵ_0 by $1/4\pi$ in the mks equations in which ϵ_0 appears (so that $1/4\pi\epsilon_0$ goes to 1) and to measure length and force in centimeters and dynes, in order to get the equivalent expression in esu. Thus the electric field in esu is given in dynes per esu or dynes per statcoulomb. Potential difference is in ergs per statcoulomb. An alternative expression for electric field is statvolts per centimeter. The ratios between these derived quantities in the mks and esu systems are

obtained as explained above. We work out a few examples, but merely list the results for other cases.

The electric field is given in both systems by

$$E = \frac{F}{q} \tag{15.7}$$

In the mks system this is in newtons per coulomb, and in the cgs system it is in dynes per statcoulomb. A given electric field in the mks system is given by

$$\begin{aligned} E_{\text{mks}} &= \frac{F \text{ (newtons)}}{q \text{ (coulombs)}} = \frac{F \text{ (dynes)}/10^5}{q \text{ (statcoulombs)}/3 \times 10^9} \\ &= E_{\text{esu}} \times 3 \times 10^4 \end{aligned} \tag{15.8}$$

Thus the field is given by fewer mks than esu units, so the esu is 3×10^4 larger than the mks unit.

For potential we may use

$$\begin{aligned} V_{\text{mks}} \text{ (volts)} &= \frac{\text{work (joules = newton-m)}}{\text{charge (coulomb)}} \\ &= \frac{\text{dyne-cm}/10^7}{\text{statcoulomb}/3 \times 10^9} = 300 V_{\text{esu}} \end{aligned} \tag{15.9}$$

The esu unit of voltage is called the *statvolt*. It is larger than the volt by a factor of 300.

Another way to arrive at this result is to use the equation

$$V_{\text{mks}} = \frac{1}{4\pi\epsilon_0} \Sigma \frac{q \text{ (coulombs)}}{r \text{ (meters)}}$$

which becomes

$$V_{\text{esu}} = \Sigma \frac{q \text{ (statcoulombs)}}{r \text{ (cm)}}$$

in the esu system. Use of these equations gives the same ratio as above between the units of the two systems.

Similarly, as given in Example 3.5*d*, the potential difference between two points in a field due to a line of charges,

$$V_{AB} \text{ (volts)} = \frac{\sigma}{2\pi\epsilon_0} \ln \frac{r_A}{r_B}$$

becomes

$$V_{AB} \text{ (statvolts)} = 2\sigma \ln \frac{r_A}{r_B}$$

For capacitance, using $C = Q/V$, we find that in esu the unit of capacitance is the centimeter. This is also often called the *statfarad*. The relationship between units is

$$1 \text{ farad} = 10^{11} \text{ cm} \tag{15.10}$$

Equations involving ϵ_0 are different in the two systems. Thus the capacitance of parallel plates in the mks system is

$$C = \frac{A\epsilon_0}{d} \qquad \text{farads}$$

and in the esu system it is

$$C = \frac{A}{4\pi d}$$

When matter is present, ϵ_0 is replaced by ϵ in these capacitance equations in the mks system, where $\epsilon = K\epsilon_0$ and K is the dielectric constant. In the esu system we replace ϵ by $K\epsilon_0$ and replace the ϵ_0 as usual by $1/4\pi$. K has the same value in both systems.

Further differences and similarities are listed in Table 15.1.

Table 15.1

	mks *	*esu* *
Gauss' law: no differences except for 4π and ϵ_0 factors	$\int \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0}\Sigma q$	$\int \mathbf{E}\cdot d\mathbf{S} = 4\pi\Sigma q$
Relations between field quantities: slightly different definitions result in some modifications †	$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}$ $P = \epsilon_0\chi E$ $D = \epsilon E$ $\epsilon = \epsilon_0 K$ $K = 1 + \chi$	$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ $P = \chi E$ $D = KE$ $K = 1 + 4\pi\chi$
Energy stored in the electric field (unit volume)	$\frac{\epsilon_0 E^2}{2}$ joules	$\frac{E^2}{8\pi}$ ergs

* In each case, E must be measured in appropriate units.

† The absence of ϵ_0 from the esu equations eliminates the use of permittivity (ϵ), leaving only the dielectric constant K. Since the dielectric constant relates to the ratio of capacitance of a capacitor with and without a dielectric filling, it is the same in all units. As a result, the susceptibility of a dielectric per unit volume (cubic meter) in the mks system is just 4π times the susceptibility per unit volume (cubic centimeter) in esu.

15.3 The Absolute Electromagnetic System of Units (EMU)

The fundamental equation in this system involves the magnetic effects of currents. We may use the equation giving the magnetic force between two parallel current elements for defining unit current in both the mks and the emu systems. For parallel elements, Eq. (6.6) becomes

$$dF = \frac{\mu_0}{4\pi}\frac{i_1\,dl_1 i_2\,dl_2}{r_{12}} \tag{15.11}$$

where $\mu_0/4\pi = 10^{-7}$ henrys/m. This equation is modified in the emu system by replacing the constant by 1 and making the usual changes from newtons and meters to dynes and centimeters. Unit current (ampere) in the mks system is obtained by setting the quantities in Eq. (15.11) equal to unit value. When we solve for i, we find

$$i^2\ (\text{amp}^2) = \frac{4\pi}{\mu_0} F\ (\text{newtons})$$

In emu we have

$$i^2\ (\text{emu}^2) = F\ (\text{dynes})$$

Making the usual comparison, we find

$$i^2\ (\text{amp}^2) = 10^7\ F\ (\text{dynes})/10^5 = 10^2 i^2\ (\text{emu}^2)$$

or

$$10\ \text{amp} = 1\ \text{emu or abampere} \tag{15.12}$$

Thus the abampere is ten times larger than the ampere. Emu quantities are identified by the prefix *ab-* just as esu quantities have the prefix *stat-*. This same ratio holds between the coulomb and abcoulomb, since they are both related to the current through the same unit of time (second).

The magnetic induction field relationship is obtainable through the fundamental equation

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{(i\,\mathbf{dl} \times \hat{\mathbf{r}})}{r^2} \tag{6.2}$$

In the emu system the constant is again taken as 1. The mks unit of field we found to be the weber per square meter. In emu the

unit of field is the gauss. Comparison shows that

$$1 \text{ weber/m}^2 = 10^4 \text{ gauss} \tag{15.13}$$

In Table 15.2, we give some comparisons between mks and emu equations.

Table 15.2

mks	*emu* *
$B = \mu_0(H + M)$	$B = H + 4\pi M$
$\mu = \mu_0(1 + \chi)$	$\mu = 1 + 4\pi\chi$
$B = \mu H$	$B = \mu H$

* In emu, $\mu_{vac} = 1$ and $B_{vac} = H_{vac}$.

15.4 Combining ESU and EMU

There are many situations where both electrostatic and magnetic measurements are involved. An important case is the motion of a charged particle in both electric and magnetic fields. The usual procedure in this situation is to convert all quantities to the appropriate system. As a first step we obtain the relationship between the statcoulomb and the abcoulomb. We have already shown that

$$3 \times 10^9 \text{ statcoulombs} = 1 \text{ coulomb}$$

and

$$1 \text{ abcoulomb} = 10 \text{ coulombs}$$

We can write at once

$$3 \times 10^{10} \text{ statcoulombs} = 1 \text{ abcoulomb} \tag{15.14}$$

In all the relationships we have found, where the factor 3×10^{10} appears, this is the velocity of light c, and its correct value is $(2.997930 \pm 0.000003) \times 10^{10}$ cm/sec. We have used 3×10^{10} for this number, though corrections are easily made if necessary.

In the mks system, the force in newtons on a charge moving in a combined E and B field is given by

$$\mathbf{F} = e\mathbf{E} + e(\mathbf{v} \times \mathbf{B}) \tag{15.15}$$

using mks units for all quantities. This equation is correct in the cgs system when esu is used for the first e and emu for the second e, using E in esu and B in emu.

15.5 Conversion Table

To convert a quantity that has been calculated or measured in practical (or mks) units to esu (or emu), multiply it by the corresponding factor given in the appropriate column of Table 15.3.

Table 15.3

Quantity	*Practical (or mks) unit*	*Factor*	
		esu	*emu*
Charge	**Coulomb**	3×10^{9}	10^{-1}
Current	**Ampere**	3×10^{9}	10^{-1}
Potential	**Volt**	$1/300$	10^{8}
Power	**Watt**	10^{7}	10^{7}
Resistance	**Ohm**	$1/(9 \times 10^{11})$	10^{9}
Inductance	**Henry**	$1/(9 \times 10^{11})$	10^{9}
Capacitance	**Farad**	9×10^{11}	10^{-9}

Thus if a measured quantity is 1 volt, this is 1/300 of the (larger) esu units of voltage (1/300 statvolt), or 10^8 emu units (or abvolts). Similarly, if a calculated quantity were, say, 20 statvolts, this would be 6,000 volts.

The mks and cgs systems represent the major classifications of unit systems, but it must be pointed out that there are a number of other minor variants in the literature. The mks system we have used is called the rationalized mks system. The unrationalized mks system is obtained from the rationalized system by making the following substitutions:

$4\pi\epsilon_0$ is replaced by ϵ_0

$4\pi\epsilon$ is replaced by ϵ

$\dfrac{\mu_0}{4\pi}$ is replaced by μ_0

$\dfrac{\mu}{4\pi}$ is replaced by μ

D is replaced by $\dfrac{D}{4\pi}$

H is replaced by $\dfrac{H}{4\pi}$

In addition, the definitions of some of the derived quantities such as χ, D, and P, M, and H may be different by factors of μ_0 or ϵ_0 in some literature.

APPENDIX A

Poisson's and Laplace's Equations

In the most general situation, where the electric field in a region may be due both to the presence of distant charges and to a distribution of charges in the region, Gauss' law gives certain limits to the way in which the potential can vary from point to point. We derive the expression for these limitations here and give an example of its application to a particular problem. Let the charge density at every point in the region be given by ρ coulombs/m^3, where ρ has a particular value for every value of x, y, and z of a rectangular-coordinate system. We now apply Gauss' law to an element of volume and derive from this a differential expression for the limitations on V.

Figure A.1 shows an element of volume $dx\,dy\,dz$, whose center is at $x = a$, $y = b$, and $z = c$. When we apply Gauss' law, we have

$$\int_{CS} \mathbf{E}\cdot d\mathbf{S} = \frac{\rho}{\epsilon_0}\,dx\,dy\,dz$$

We can evaluate the Gaussian integral $\int_{CS} \mathbf{E}\cdot d\mathbf{S}$ over the volume $dx\,dy\,dz$ by taking the sum of the normal components of E times the surface areas of each of the six cube sides. That is, we compute the flux of E through each cube side. We examine first the two sides in the yz plane. The x coordinates of these two faces are $x_1 = a - \frac{1}{2}\,dx$ and $x_2 = a + \frac{1}{2}\,dx$. The x component of E is $E_x = -\partial V/\partial x$. We may now write the Gaussian integral for the flux through these two

sides. The left-hand face contributes $dy\,dz\,(\partial V/\partial x)_{a-\frac{1}{2}dx}$ and the right-hand face contributes $-dy\,dz\,(\partial V/\partial x)_{a+\frac{1}{2}dx}$. Here the subscripts mean that the $\partial V/\partial x$ is to be evaluated at the two faces. The sum of these two terms is the *net* flux through the *yz* faces. This may be written as

$$-\left[\left(\frac{\partial V}{\partial x}\right)_{a+\frac{1}{2}dx} - \left(\frac{\partial V}{\partial x}\right)_{a-\frac{1}{2}dx}\right] dy\,dz = -\frac{\partial^2 V}{\partial x^2}\,dx\,dy\,dz$$

This relationship results from the fact that the difference between the values of $\partial V/\partial x$ at the two faces is just the rate of change of

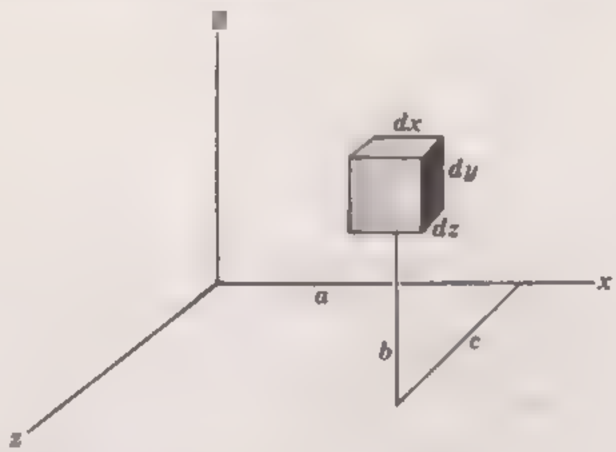


***Fig. A.1** Elementary volume used in obtaining Poisson's and Laplace's equations from Gauss' law.*

$\partial V/\partial x$ along x (i.e., $\partial^2 V/\partial x^2$ times the separation dx between the two positions).

Similar reasoning can be applied to the other two pairs of cube faces, and we get from Gauss' law:

$$\left(\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}\right) dx\,dy\,dz = -\frac{\rho}{\epsilon_0}\,dx\,dy\,dz$$

Since this relation holds for any small volume $dx\,dy\,dz$, it follows that

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{\rho}{\epsilon_0} \qquad \text{(A.1)}$$

Thus the inverse-square law puts this limitation on the second derivatives of V with respect to x, y, and z. Vector notation for this quantity is $\nabla^2 V$, called *del squared V*. Any static electric field must satisfy this relationship, which we rewrite in vector notation as

$$\nabla^2 V = -\frac{\rho}{\epsilon_0} \qquad \text{(A.2)}$$

Some insight can be given into the meaning of this by taking the important case of a region in which there is no net charge. Then Poisson's equation becomes the equation of Laplace,

$$\nabla^2 V = 0 \tag{A.3}$$

Suppose the field configuration is such that the flux of E through the volume is zero except through the $dx\,dy$ faces. The equation then tells us that the value of E is the same at both faces or that the net flux through the volume is zero. But this is just what we know from Gauss' law or from the idea that lines of force are continuous and originate and end only on charges. The field we have just described is one that is uniform and in the x direction.

APPENDIX B

Conducting Sphere in a Uniform Field

We examine here, as an example of the general electrostatic boundary-value problem, the case of an uncharged conducting sphere placed in an otherwise uniform field.

A useful way of formulating the problem is to look for a solution having the following attributes: (1) It obeys Poisson's or Laplace's equation, depending on whether or not there is space charge. We limit ourselves to a case involving conductors in a vacuum, so $\nabla^2 V = 0$ (see Appendix A). (2) It gives equipotential surfaces as required by the problem. (3) It behaves appropriately at infinity.

According to the uniqueness theorem,[1] which we are stating without proof, we have a guarantee that any solution that does all these things is the only solution. As shown in Fig. B.1, we let the field far from the sphere be E_0 (this is also the field in the absence of the sphere) and the radius of the sphere be a, and we assume the net charge on the sphere to be zero. There is an equipotential plane cutting through the middle of the sphere and perpendicular to the electric field direction. We define this plane to be at zero potential. The problem now is to find an expression for the potential such that $\nabla^2 V = 0$, the sphere is an equipotential surface, and at points far away the potential is the same as in the absence of the perturbing effects of the sphere, i.e., the potential consistent with a uniform

[1] A proof of the uniqueness theorem for static electric and magnetic fields is given, for example, in W. V. Houston, "Principles of Mathematical Physics," 2d ed., McGraw-Hill Book Company, Inc., 1948, pp. 271–273.

field. Using cylindrical coordinates for convenience, at points far away from the sphere, $V = -E_0x$, or $V = -E_0r\cos\theta$. Thus, whatever our final expression for V, it must reduce to this for large r. Another consideration is that the surface of the sphere must be at a constant potential of zero. Thus an added term must also contain $\cos\theta$, so that for all positions on the sphere, the sum of the terms can be zero. However, for large r this second term must approach zero. An expression to try is one containing $\cos\theta/r^2$. We try $V = -E_0r\cos\theta + (A/r^2)\cos\theta$, where A is a constant to be determined. Both terms are *harmonic functions*, all of which are solutions of

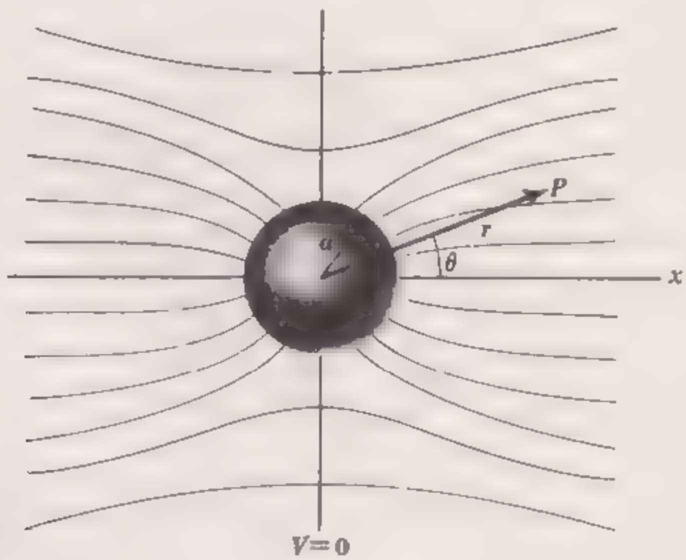


Fig. B.1 Conducting sphere in a uniform external field.

$\nabla^2V = 0$. When we set $V = 0$ for $r = a$, to satisfy the boundary condition at the surface of the sphere, we find $A = E_0a^3$, so the expression becomes $V = -E_0r\cos\theta + (E_0a^3/r^2)\cos\theta$. This expression is valid for all $r > a$. Boundary conditions are satisfied if we use for the second term in the potential $(E_0a^2/r)\cos\theta$. However, this term would not satisfy Laplace's equation and therefore is ruled out. Discussion and development of the possible forms of solutions to Laplace's equation are found in more advanced treatments of this subject. Since this expression fits the boundary conditions at $r = a$ and at $r = \infty$ and satisfies Laplace's equation, it must be the solution to the problem.

The second term, $(E_0a^3/r^2)\cos\theta$, amounts to the perturbation of the original uniform field by the conducting sphere. It corresponds to the effect of a dipole of dipole moment $p = E_0a^3$ placed at the position of the center of the sphere, with its axis along x.

The field at the surface of the sphere is obtained by taking the appropriate derivative of V. Since the field is normal to a conducting surface, it is along r for a conducting sphere. Thus,

$$E = -\frac{\partial V}{\partial r} = E_0 \cos\theta + \left(\frac{2E_0a^3}{r^3}\right)_{r=a} \cos\theta = 3E_0 \cos\theta$$

Along the x axis the field at the surface is $3E_0$. At points on the sphere where $\theta = 90°$, both V and E are zero.

APPENDIX C

Dielectric Sphere in a Uniform Field

Here we discuss, partly without proof, the way in which a dielectric body perturbs a uniform field in which it is placed. We begin with the study of a dielectric sphere placed in a uniform field $\mathbf{E}_0$ as shown in Fig. C.1. This is related to the problem of the spherical conductor

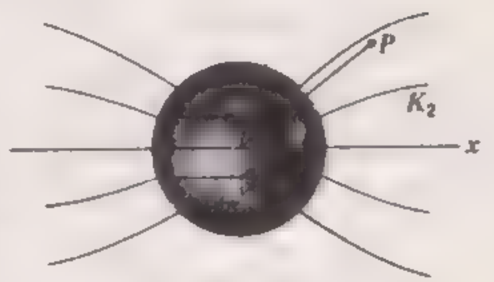


Fig. C.1 Dielectric sphere inserted in a uniform electric field. Lines of D are shown.

in a uniform field as discussed in Appendix B. However, the boundary condition at the surface of the sphere is now different, and the field inside the dielectric sphere is not zero as it was in the conducting sphere. We keep the problem general by assigning dielectric constant K_1 to the sphere and K_2 to the region outside the sphere. The potential functions are V_1 inside the sphere ($r \leq a$) and V_2 outside ($r \geq a$). The solution to this problem must satisfy the following requirements:

1. $\nabla^2 V_1 = 0$ and $\nabla^2 V_2 = 0$; that is, both inside and outside, Laplace's equation must be satisfied. As in the case of the conducting sphere studied in Appendix B, the potential function required is made up of harmonic functions.

2. $V_2 + E_0 r \cos\theta$ must remain finite at infinity. The original uniform field has a potential given by $-E_0 x$ or $-E_0 r \cos\theta$ when we assume zero potential at the center of the sphere. V_1 must therefore contain this negative term. The requirement is then simply that the additional term or terms due to the presence of the sphere must become negligible at large distances.

3. V_1 must remain finite for all $r \leq a$ (inside the sphere).

4. $V_2 = V_1$ for $r = a$ at all angles θ. If this were not true, V would be a step function at the surface. This sudden change in V would require the rate of change of V to be infinite at the discontinuity. This would require infinite field.

5. $K_1 \partial V_1/\partial r = K_2 \partial V_2/\partial r$ at $r = a$ for all θ. This is simply the boundary condition $D_{n1} = D_{n2}$, which can be written $K_1 E_{n1} = K_2 E_{n2}$.

On the basis of our earlier result for the conducting sphere, we try the following solutions, which are made up of harmonic functions and which therefore are guaranteed to satisfy Laplace's equation.

$$V_2 = -E_0 r \cos\theta + \frac{A}{r^2}\cos\theta \tag{C.1}$$

$$V_1 = Br\cos\theta \tag{C.2}$$

Even without specifying the constants A and B, these solutions satisfy requirements 1 to 3. In order to satisfy (4) and keep V single-valued at the surface, we set

$$-E_0 a + \frac{A}{a^2} = Ba \tag{C.3}$$

In order to satisfy (5), we take $K_1 \dfrac{\partial V_1}{\partial r} = K_2 \dfrac{\partial V_2}{\partial r}$, to get

$$K_2\left(-E_0 - \frac{2A}{a^3}\right) = K_1 B \tag{C.4}$$

If we solve (C.3) and (C.4) for A and B, we find

$$A = \frac{K_1 - K_2}{K_1 + 2K_2} E_0 a^3$$

$$B = \frac{-3K_2}{K_1 + 2K_2} E_0$$

For a sphere in a vacuum we put $K_2 = 1$, and for a spherical cavity in a dielectric we put $K_1 = 1$. For all cases, the potential inside the

sphere can be written as $V_1 = Bx$. This means that the field inside the sphere is uniform. For the dielectric sphere in a vacuum, the original uniform field E_0 has been reduced to a value $[3/(K_1 + 2)]E_0$. This reduction is due to the uniform field of the polarization charges on the surface of the sphere acting in opposition to the original field. The size of this depolarization field E_{dep}, which acts in opposition to the original field E_0, is given for this case by

$$E_{dep} = E_0 - E_2 = \left(1 - \frac{3}{K_1 + 2}\right) E_0 = \left(\frac{K_1 - 1}{K_1 + 2}\right) E_0 \tag{C.5}$$

This opposing field may be expressed in terms of a *depolarizing factor* L relating the field to the state of polarization of the dielectric,

$$\epsilon_0 E_{dep} = LP \tag{C.6}$$

ϵ_0 is included in order to satisfy dimensional requirements.

The value of L for a sphere is obtained when we note that

$$P = \epsilon_0(K - 1)E \tag{C.7}$$

where the subscript on K is omitted for convenience, and where E is the macroscopic field inside the dielectric. This equation follows from (5.3) and (5.9). Substitution in Eq. (C.6) gives

$$L = \frac{(K - 1)/(K + 2)}{3(K - 1)/(K + 2)} = \frac{1}{3} \tag{C.8}$$

In this case the potential V_2 outside the dielectric sphere consists of a term $-E_0 r \cos\theta$ from the original unperturbed uniform field, plus a term $(A/r^2)\cos\theta$. This latter term gives rise to a dipole field, the same as we found for a dipole in Example 3.7*b*. Thus the field outside a uniformly polarized dielectric sphere in an originally uniform field is just the original uniform field, superposed on which is a dipole field due to the polarized sphere.

For convenience we have drawn lines of **D** rather than lines of **E** in Fig. C.1, since lines of **D** are continuous across the boundary of the dielectric, while lines of **E** are not. Although **D** inside the dielectric sphere is increased over the value in the original uniform field, **E** is decreased, as a result of the depolarization field.

In choosing a sphere, we have taken one of the few shapes to which a depolarization factor accurately applies, since in general

the field inside a dielectric body immersed in a uniform field is not uniform inside the body. Using isotropic material, only specimens having the shape of a general ellipsoid have uniform internal fields. If one of the principal axes of the ellipsoid is parallel to the applied field, the polarization and the depolarizing field are along the same line as the original external field (assuming an isotropic dielectric).

APPENDIX D

Complex Numbers

For the benefit of those students who have been introduced to the methods of complex numbers, we discuss briefly their use in solving a-c circuit problems. Complex notation for these problems is very attractive since it greatly simplifies their solution.

Complex notation is particularly indicated whenever sine or cosine functions are involved. We show this by considering the two sinusoidal functions that can be used for the instantaneous voltage of an a-c source, either

$$v = V_0 \cos \omega t \tag{10.2}$$

or

$$v = V_0 \sin \omega t \tag{10.1}$$

A-c problems can be solved using either of these functions, and the physical results will be identical since the only difference between the two is their value at $t = 0$. It requires only a shifting of the time axis by $\pi/2$, or ¼ cycle, to convert one expression into the other. [Thus $\cos(\omega t + \pi/2) = \sin \omega t$.]

In complex notation both functions are included in the single expression

$$v = V_0 e^{j\omega t} \tag{D.1}$$

where $j = (-1)^{1/2}$. To obtain this result, we need merely to write the series expansion for the exponential e^x,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \tag{D.2}$$

and note that

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

and

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

are series expansions for $\sin x$ and $\cos x$. We then substitute $j\omega t$ for x [note that $(-1)^{1/2}(-1)^{1/2} = -1$], to obtain

$$e^{j\omega t} = \cos \omega t + j \sin \omega t \tag{D.3}$$

$e^{j\omega t}$ is a *complex* number because it is the sum of two terms, one *real* and the other *imaginary*, that is, containing j.

It is useful to consider complex numbers as plotted in the complex plane, shown in Fig. D.1. Here the complex number $\mathbf{N} = \mathbf{A} + j\mathbf{B}$ is plotted: the real part $\mathbf{A}$, along the *real* (horizontal) axis, the imaginary part $\mathbf{B}$, along the imaginary (vertical) axis. The complex number N is given by the vector sum as shown. Knowledge of $\mathbf{A}$ and $\mathbf{B}$ is equivalent to a knowledge of the magnitude of $\mathbf{N}$ and the angle θ.

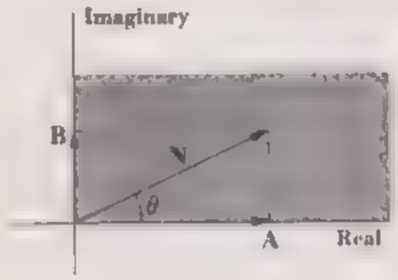


***Fig. D.1** A complex number N plotted in the complex plane, showing real and imaginary components A and B.*

Let us now use the complex notation for obtaining the solution to the series LCR circuit. We write the differential equation that relates the sinusoidal generator voltage to the sum of the instantaneous voltages across R, C, and L:

$$V = V_0 e^{j(\omega t - \theta)} = iR + L\frac{di}{dt} \pm \frac{1}{C}\int i\, dt \tag{D.4}$$

The angle θ is included in the exponent to allow for the phase angle between current and voltage. Since the current i is sinusoidal, we can write

$$i = I_0 e^{j\omega t} = I_0 (\cos \omega t + j \sin \omega t) \tag{D.5}$$

At this stage we need not commit ourselves to the real or imaginary part but can carry both parts. At the end we can pick either the real part (appropriate to $i = I_0 \cos \omega t$) or the imaginary part (appropriate to $i = I_0 \sin \omega t$). Next we obtain di/dt and $\int i\,dt$ as required above:

$$\frac{di}{dt} = I_0 \frac{d}{dt} e^{j\omega t} = I_0 j\omega e^{j\omega t} \tag{D.6}$$

$$\int i\,dt = I_0 \int e^{j\omega t}\,dt = I_0 \frac{1}{j\omega} e^{j\omega t} \tag{D.7}$$

Substitution in Eq. (D.4) gives

$$\begin{aligned} V_0 e^{j(\omega t - \theta)} = V = I_0 \left(R e^{j\omega t} + j\omega L e^{j\omega t} + \frac{1}{j\omega C} e^{j\omega t} \right) \\ = \left(R + j\omega L + \frac{1}{j\omega C} \right) I_0 e^{j\omega t} \end{aligned}$$

When we take, say, the real part of this equation, we find

$$i = \frac{V_0 \cos (\omega t - \theta)}{[R + j(\omega L - 1/\omega C)]} \tag{D.8}$$

essentially as in Eq. (10.16). The complex expression in the denominator has the same meaning as the impedance **Z** obtained earlier. This may be seen by comparison with Fig. 10.12, if we associate the real part R with the x axis in that figure and the imaginary part $j(\omega L - 1/\omega C)$ with the y axis. It is also apparent that the phase angle is given, as in our earlier treatment, by

$$\tan \theta = \frac{\omega L - 1/\omega C}{R} \tag{10.17}$$

We have thus seen that the a-c series LCR circuit is very easily handled if we write the three impedance terms as R, $j\omega L$, and $1/j\omega C$. The parallel LCR circuit can be solved in exactly the same way. We leave to the student the problem of showing that the parallel LCR circuit equation (10.26) can be obtained by using complex notation.

APPENDIX E

Effect of Primary and Secondary Multiplication Processes on Electric Conductivity in Gases

Figure E.1 shows parallel-plate electrodes in a gas volume. We consider the following processes. An external source of radiation produces n_0 electrons per second very close to the cathode. Additional n_s electrons per second are emitted from the cathode by secondary processes resulting from the primary multiplication process in the gas. These may be both photoelectrons and electrons resulting from positive-ion collisions at the cathode. The total flow of electrons from the cathode region, $(n_0 + n_s)$/sec, is then multiplied by the avalanche process and results in n electrons per second reaching the

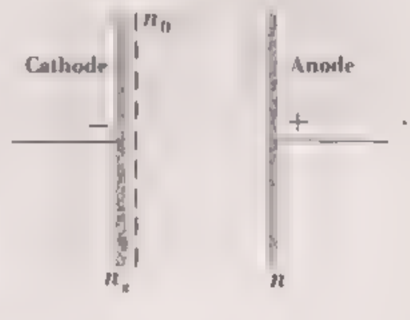


***Fig. E.1** Multiplication of electrons in a gas discharge. The number of electrons formed near the cathode by an external source is n_0, the enhanced number arriving at the anode is n, and the secondary electrons produced at the cathode is n_s. Primary and secondary multiplication processes together can allow self-maintaining discharge.*

anode. We can write

$$n = (n_0 + n_s)e^{\alpha x} \tag{E.1}$$

using Eq. (13.4). We now define a quantity γ as the probability that a secondary electron will be freed at the cathode by a secondary process associated with each primary multiplication event. Now the number of primary processes per second that take place is just the difference between the number of electrons per second arriving at the anode and the number per second starting at the cathode. This is

$$n - (n_0 + n_s)$$

The number of secondary electrons per second produced at the cathode is then given by

$$n_s = \gamma[n - (n_0 + n_s)]$$

From this last equation, with a little rearranging, we can write the expression for the total number of electrons per second leaving the cathode,

$$n_s + n_0 = \frac{\gamma n + n_0}{1 + \gamma}$$

When we substitute this in Eq. (E.1) and solve for n, we find

$$n = n_0 \frac{e^{\alpha x}}{1 - \gamma(e^{\alpha x} - 1)}$$

Now $e^{\alpha x}$ is normally much greater than 1, so we can use the approximate equation

$$n = n_0 \frac{e^{\alpha x}}{1 - \gamma e^{\alpha x}} \tag{E.2}$$

Since the current is proportional to the number of electrons per second, this can be written as

$$i = i_0 \frac{e^{\alpha x}}{1 - \gamma e^{\alpha x}} \tag{E.3}$$

Some Important Physical Constants

Electron charge	e	1.60×10^{-19} coulomb
Permittivity constant	ϵ_0	8.85×10^{-12} coulomb^2/newton-m^2
	$\dfrac{1}{4\pi\epsilon_0}$	9×10^9 newton-m^2/coulomb^2
Permeability constant	μ_0	$1.26 \times 10^{-6} = 4\pi \times 10^{-7}$ weber/amp-m
	$\dfrac{\mu_0}{4\pi}$	10^{-7} weber/amp-m
Magnetic induction field	B	1 weber/m^2 = 10,000 gauss
Speed of light	C	3.00×10^8 m/sec = 1.86×10^5 miles/sec
Electron rest mass	m_e	9.11×10^{-31} kg
Proton rest mass	m_p	1.67×10^{-27} kg
Electron charge to mass ratio	$\dfrac{e}{m_e}$	1.76×10^{11} coulombs/kg
Magnetic moment of electron	μ_e	9.29×10^{-18} joule-m^2/weber
Planck's constant	h	6.63×10^{-34} joule-sec
Boltzman's constant	k	1.38×10^{-23} joule/°K
Electron volt	eV	1.60×10^{-19} joule

Answers to Odd-numbered Problems

1.1 $Q_2 = -4Q_1$

1.3 (a) $F = 0$

(b) $F = 9 \times 10^{13} Q^2$ newtons

1.5 (a) $F = 0$

(b) $F = 2.88 \times 10^7$ newtons

1.7 4.27×10^{42}

2.1 (a) $E = 0$

(b) $E = \dfrac{4Q}{4\pi\epsilon_0 a^2}\left(\dfrac{3}{2}\right)^{-3/2}$

$= 2.2 \times 10^{12} Q$ newtons/coulomb

(c) $E = \dfrac{4Q}{4\pi\epsilon_0}\dfrac{1}{0.03}$ newtons/coulomb

2.3 -1.77×10^{-8} coulomb

2.5 $\frac{9}{4}\mu \times 10^9$ newtons/coulomb

2.7 (a) $E = 0$

(b) $E = \dfrac{12.5Q}{4\pi\epsilon_0}$ newtons/coulomb

(c) $E = \dfrac{25Q}{4\pi\epsilon_0}$ newtons/coulomb

(d) $E = \dfrac{4Q}{4\pi\epsilon_0}$ newtons/coulomb

2.9 $E = \dfrac{1}{4\pi\epsilon_0}\dfrac{2Q}{\pi r^2}$ newtons/coulomb

2.11 0

2.13 $F = KaQ = Kp$ newtons

2.15 $E = \dfrac{Q}{4\pi\epsilon_0 2R^2}$ newtons/coulomb

3.1 (a) $V_s = \dfrac{p}{4\pi\epsilon_0(a^2 - l^2/4)}$ volts

$W = QV_s$ volts

(b) $V_s = \dfrac{p}{4\pi\epsilon_0 a^2}$ volts

$E = \dfrac{2p}{4\pi\epsilon_0 a^3}$ volts

3.3 $V_{in} = \dfrac{\rho}{6\epsilon_0}(3R^2 - r^2)$ volts

$V_{out} = \dfrac{\rho R^3}{3\epsilon_0 r}$ volts

$E_{in} = \dfrac{\rho r}{3\epsilon_0}$ volts/m

$E_{out} = \dfrac{\rho R^3}{3\epsilon_0 r^2}$ volts/m

3.5 $V = 10^7$ volts

3.7 (a) $Q_{out} = 4\pi\epsilon_0 V_1 b$

$Q_{in} = -Q_1$

(b) $r < a$: $E = 0$

$V = V_1 + \dfrac{Q_1(b-a)}{4\pi\epsilon_0 ab}$ volts

$a < r < b$: $E = \dfrac{Q_1}{4\pi\epsilon_0 r^2}$ volts/m

$V = V_1 + \dfrac{Q_1(b-r)}{4\pi\epsilon_0 rb}$ volts

$r > b$: $E = \dfrac{V_1 b}{r^2}$ volts/m

$V = \dfrac{V_1 b}{r}$ volts

3.9 $V_{12} = \dfrac{Q}{2\pi\epsilon_0}\ln\dfrac{r_2}{r_1}$ volts

3.11 $V = 1.87 \times 10^7$ m/sec

3.13 $V_2 = 2^{3/5} \cdot V_1$ volts

3.15 (b) $x = 32.3$ cm, -7.65 cm, -20 cm
(c) Equilibrium at $x = -20$ cm, stable for $-$ charge, unstable for $+$ charge

3.17 $E_r = \dfrac{2a\cos\theta}{r^3} + \dfrac{b}{r^2}$ volts/m

$E_\theta = \dfrac{a\sin\theta}{r^3}$ volts/m

4.1 $C = \dfrac{2\pi\epsilon_0 l}{\ln(b/a)}$ farads

4.3 $C = \dfrac{(C_1 + C_2)C_3}{C_1 + C_2 + C_3}$ farads

4.5 $C_2 = \dfrac{C_1 d}{d - a}$ farads

4.7 $F = \dfrac{A\epsilon_0 V^2}{2d^2}$ newtons

4.9 $C = \dfrac{6\epsilon_0 A}{d}$ farads

4.11 $V_1 = 150$ volts
$V_2 = 50$ volts
$Q_1 = Q_2 = 3 \times 10^{-4}$ coulomb

4.13 (a) $U = \dfrac{Q_2}{8\pi\epsilon_0 R}$ joules
(b) $r = 2R$

4.15 $Q = -q$ coulomb

4.17 $\Delta U = \dfrac{1}{2} Q_0{}^2 \dfrac{C_2}{C_1{}^2 + C_1 C_2}$ joules

5.1 $\sigma_1 = -P$
$\sigma_2 = 0$
$\sigma_3 = P\sin\theta$

5.3 $V_2 = \dfrac{V_1}{K}$ volts

$U_2 = \dfrac{U_1}{K}$ joules

Slab pulled in

5.7 $P = -\sigma_f$
$\chi = \infty$

5.9 $V_2 = \dfrac{V_1}{1 + K}$ volts

$\Delta Q = \dfrac{CV_1 K}{1 + K}$ coulombs

5.11 $\Phi_2 = 24°28'$

6.1 $\dfrac{F}{l} = \dfrac{\mu_0 i i'}{2\pi a} = 2 \times 10^{-4}$ newton/m (attractive)

6.3 $B = \dfrac{\mu_0 j r}{2}$ webers/m² $\quad r < a$

$B = \dfrac{\mu_0 a^2 j}{2r}$ webers/m² $\quad r > a$

6.5 $B = \dfrac{\mu_0 i}{4r}$ webers/m²

6.7 $B = \dfrac{\mu_0 i r}{2\pi a^2}$ webers/m² $\quad r < a$

$B = \dfrac{\mu_0 i}{2\pi r}$ webers/m² $\quad a < r < b$

$B = \dfrac{\mu_0 i}{2\pi r}\left(1 - \dfrac{r^2 - b^2}{c^2 - b^2}\right)$ webers/m² $\quad b < r < c$

$B = 0 \quad r > c$

6.9 (a) $j^s = \dfrac{Ni}{L} = 3 \times 10^4$ amp/m
(b) $B_1 = \dfrac{\mu_0 Ni}{L} = 12\pi \times 10^{-3}$ weber/m²
(c) $B_2 = \dfrac{\mu_0 Ni}{2L} = 6\pi \times 10^{-3}$ weber/m²
(d) $\Phi = B_1 A = 48\pi^2 \times 10^{-7}$ weber
(e) $\Phi = B_2 A = 24\pi^2 \times 10^{-7}$ weber

6.11 $B = \dfrac{\mu_0 \sigma \omega a^4}{8b^3}$ webers/m² $\quad b \gg a$

6.13 $\tau = 5\pi \times 10^{-2} B \sin\theta$ newton-m

6.15 $\Phi = 1.1 \times 10^{-7}$ weber

6.17 $R_H = 6.25 \times 10^{-10}$
$N = 10^{28}$ electrons/m³

6.19 (a) $\omega = 4.8 \times 10^7$ radians/sec
(b) $W = 1.92 \times 10^{-12}$ joule $= 1.2 \times 10^7$ ev
(c) $n = 300$ revolutions

6.23 $m = 7.2 \times 10^{-26}$ Kg

6.27 $B_b = \dfrac{\mu_0 i}{\pi r}$ webers/m²

$B_a = \dfrac{\mu_0 i}{r}\left(\dfrac{1}{4} + \dfrac{1}{2\pi}\right)$ webers/m²

6.29 $B = \dfrac{\mu_0 i}{2r}$ webers/m²

7.1 $i = \frac{2}{3}\pi j_0 a^3$ amp

7.3 $v = 0.0625$ m/sec

7.5 $R = 1.7 \times 10^{-5}$ ohm

7.7 $R_2 = 1$ ohm

7.9 $R = \frac{1}{4} \times 10^{-4}$ ohm

7.11 $R_0 = (1 + \sqrt{3})R$ ohms

7.13 (a) $i = 2$ amp
(b) $i_1 = 2$ amp
$i_4 = 1$ amp
$i_8 = 1$ amp
$i_6 = 2/3$ amp
$i_{12} = 1/3$ amp
(c) $P_1 = 4$ watts
$P_4 = 4$ watts
$P_6 = 8/3$ watts
$P_8 = 8$ watts
$P_{12} = 4/3$ watts
$P_i = 4$ watts
(d) $P = 24$ watts
(e) $V = 10$ volts

7.17 (a) $R_v = 9{,}990$ ohms
(b) $R_a = 1.111$ ohms

7.19 $R_1 = 0.0278$ ohms
$R_2 = 0.250$ ohms
$R_3 = 2.50$ ohms

8.1 (a) $F = evB$ newtons upward on electrons
(b) $E = vB$ volts/m upward
(c) $V = vBL$ volts
(d) $i = \dfrac{vBL}{R}$ amp
(e) $\mathcal{E} = vBL$ volts

8.3 $L' = 3L$ henrys

8.5 $\mathcal{E} = \frac{1}{2}\omega L^2 B$ volts

8.7 $\mathcal{E}_{max} = \pi^2 f R^2 B$ volts
$i_{max} = 10^{-3}\pi^2 f R^2 B$ amp
freq $= f$

8.11 $\mathcal{E} = \dfrac{\mu_0 iab}{2\pi}\,\dfrac{V}{l(l+a)}$ volts

8.13 $L = \dfrac{\mu_0 N^2 A}{l}$ henrys

8.15 $\Phi = 0.5 \times 10^{-4}$ weber

8.17 $W = 10^4$ joules

9.1 (a) $\overset{s}{j}_{free} = 13{,}333$ amp/m
(b) $H = 13{,}333$ amp/m
(c) $\mu = 1.28 \times 10^{-6}$ weber/amp-m
(d) $M = 267$ amp/m
(e) $B = 17.1 \times 10^{-3}$ weber/m^2

9.3 (a) $L = 1.8$ henry
(b) $H = 2.39 \times 10^4$ amp/m
(c) $\overset{s}{j}_{free} = 2.39 \times 10^4$ amp/m
(d) $M = 2.39 \times 10^7$ amp/m
(e) $\overset{s}{j}_{mag} = 2.39 \times 10^7$ amp/m
(f) $U = 22.5$ joules

9.5 $\dfrac{X_1}{X_2} = (2 \cot\theta)^{-1/2}$

9.7 $\chi \doteq 3.28 \times 10^{-3}$
$p_m = 2.61 \times 10^{-3}$ amp-m^2
$M = 1.8 \times 10^6$ amp/m
$p_m = 18$ amp-m^2
$\tau = 1.8 \times 10^{-6}$ newton-m
$q_m = 180$ unit poles

9.9 $F = \dfrac{N^2 i^2 A}{\mu_0}\left(\dfrac{l}{\mu} + \dfrac{2x}{\mu_0}\right)^2$ joules

10.1 (a) $i = \dfrac{V_0}{(R^2 + \omega^2 L^2)^{1/2}} \sin(\omega t + \Phi)$

$\Phi = \tan^{-1}\dfrac{\omega L}{R}$

(b) $\Delta\Phi = 0$
(c) $\Delta\Phi = -90°$ V_R lags V_L
(d) $V_R = \dfrac{V_0 R}{(R^2 + \omega^2 L^2)^{1/2}}$ volts

$V_L = \dfrac{V_0 \omega L}{(R^2 + \omega^2 L^2)^{1/2}}$ volts

10.3 (a) $X_L = 755$ ohms
(b) $X_c = 53$ ohms
(c) $f = 15.9$ cycles/sec

10.5 (a) $V_{rms} = 85$ volts
(b) $I_0 = 6$ amp
(c) $I_{av} = 0$
(d) $I_{rms} = 4.25$ amp
(e) $p = 360$ watts

10.7 (a) $Z = 800$ ohms
(b) $P = 4$ watts
(c) $C = 3.84$ μf

10.9 $V_{rms} = 38$ volts

10.11 (a) $\omega = \dfrac{1}{(LC)^{1/2}}$ radians/sec
(b) $\omega_{max} = 0, \infty$

$\omega_{min} = \dfrac{1}{(LC)^{1/2}}$ radians/sec

(c) $\omega = \dfrac{\sqrt{3}/R \pm (3/R^2 + 4C/L)^{1/2}}{2C}$ radians/sec

10.15 $R_1 = 500$ ohms

$\dfrac{di}{dt} = -120$ amp/sec

10.17 $\dfrac{V_1}{V_2} = \dfrac{5n_1}{2n_2}$

11.1 (b) $\Delta V_p = 25$ volts

12.5 (a) $\lambda = 300$ m

(b) $E_0 = 122.6$ volts/m

$H_0 = 0.325$ amp/m

12.7 $E_0 = 1{,}020$ volts/m

$B_0 = 3.4 \times 10^{-6}$ weber/m^2

12.9 $\lambda_g = 2.04$ cm

$V_p = 4.9 \times 10^8$ m/sec

$V_g = 1.83 \times 10^8$ m/sec

Index